QC
1
A85
v.9
1956
Science
PER

Digitized by the Internet Archive
in 2024

https://archive.org/details/australian-journal-of-physics_1956_9

AUSTRALIAN JOURNAL
OF PHYSICS

VOLUME 9

MELBOURNE
1956

AUSTRALIAN JOURNAL OF PHYSICS

Published by the Commonwealth Scientific and Industrial Research Organization in collaboration with the Institute of Physics (Australian Branch), and under the general direction of the Editorial Board (Dr. N. S. Noble (Chairman), Professor J. S. Anderson, Professor Sir Macfarlane Burnet, Professor L. H. Martin, and Professor J. G. Wood). Volumes 1 to 5 of the Australian Journal of Physics and the Australian Journal of Chemistry issued as the Australian Journal of Scientific Research, Series A: Physical Sciences.

Issued quarterly, 30/- per annum

EDITORIAL ADVISORY COMMITTEE

Chairman and Editor: Dr. N. S. Noble

Members: Dr. G. H. Briggs
Professor L. H. Martin
Professor H. C. Webster

All enquiries and manuscripts should be forwarded to:

The Editor,
Australian Journal of Physics,
Commonwealth Scientific and Industrial Research Organization,
314 Albert Street, East Melbourne, C.2, Victoria.

Printed by Australasian Medical Publishing Co. Ltd., Sydney.

CONTENTS

Number 1, March 1956

PAGE

Linear Superposition in Visco-Elasticity and Theories of Delayed Effects. By E. R. Love 1

Reciprocal Static Solutions of the Equations of the Gravitational Field. By H. A. Buchdahl 13

The Accelerated Motion of Droplets and Bubbles. By T. Pearcey and G. W. Hill 19

Growing Electric Space-Charge Waves. By J. H. Piddington 31

Formulae for the Mean Losses of Energy in Collisions of Slow Electrons Moving in Diatomic Gases. By L. G. H. Huxley 44

The Scattering of 121 keV Electrons and Positrons. By L. J. Tassie .. 54

18·3 Mc/s Radiation from Jupiter. By C. A. Shain 61

Cosmic Radio Sources Observed at 600 Mc/s. By J. H. Piddington and G. H. Trent 74

Radio Emission from Novae and Supernovae. By B. Y. Mills, A. G. Little, and K. V. Sheridan 84

The Masses of the Magellanic Clouds from Radio Observations. By F. J. Kerr and G. de Vaucouleurs 90

Lunar Variations in the Ionosphere. By R. A. Duncan 112

Measurements of Counter-Gradient Heat Flows in the Atmosphere. By A. F. Bunker 133

Short Communications

On the Dipole Resonant Mode of an Ionized Gas Column. By G. H. Keitel 144

Neutron Fluxes Underground. By K. B. Mather 147

The Fitting of a Truncated Log-Normal Curve to Daily Rainfall Data. By S. C. Das 151

Composition of the ^{7}Li (p,γ) Radiation Below the 441 keV Resonance. By J. G. Campbell 156

Ionospheric Recorders and Sporadic E. By J. A. Thomas, A. C. Svenson, and H. E. Brown 159

Processes Controlling Ionization Distribution in the F_2 Region of the Ionosphere. By D. F. Martyn 161

Number 2, June 1956

PAGE

Conduction of Heat in an Infinite Region Bounded Internally by a Circular Cylinder of a Perfect Conductor. By J. C. Jaeger 167

Thermal and Electrical Conductivities of Iron, Nickel, Titanium, and Zirconium at Low Temperatures. By W. R. G. Kemp, P. G. Klemens, and G. K. White 180

The Heat Capacity of Copper Below 4·2 °K. By J. A. Rayne . . 189

Strip Integration in Radio Astronomy. By R. N. Bracewell 198

Emission Nebulae as Radio Sources. By B. Y. Mills, A. G. Little, and K. V. Sheridan 218

Sporadic *E* at Brisbane. By J. A. Thomas 228

A Study of "Spread-*F*" Ionospheric Echoes at Night at Brisbane. I. Range Spreading (Experimental). By R. W. E. McNicol, H. C. Webster, and G. G. Bowman 247

A Study of "Spread-*F*" Ionospheric Echoes at Night at Brisbane. II. Interpretation of Range Spreading. By R. W. E. McNicol and H. C. Webster 272

The Measurement of the Thermal Conductivities of Rocks by Observations in Boreholes. By A. Beck, J. C. Jaeger, and G. Newstead 286

Corrigendum 296

Number 3, September 1956

PAGE

Two-dimensional Aerial Smoothing in Radio Astronomy. By R. N. Bracewell 297

On the Association of Solar Radio Emission and Solar Prominences. By J. P. Wild and H. Zirin 315

Ionospheric Reflections from Heights below the *E* Region. By J. B. Gregory 324

Cusp Type Anomalies in Variable Frequency Ionospheric Records. By G. H. Munro and L. H. Heisler 343

Divergence of Radio Rays in the Ionosphere. By G. H. Munro and L. H. Heisler 359

The Theory of Strong Katabatic Winds. By F. K. Ball 373

The Condition of Regularity in Simple Markov Chains. By J. Gani .. 387

Statistical Analysis of Australian Pressure Data. By S. C. Das.. .. 394

Short Communications

Scattering of High Energy Electrons by Carbon and the Nuclear Density Distribution. By L. J. Tassie 400

The Thermal Expansion of a Grossularite Garnet. By L. G. Wilson .. 403

NUMBER 4, DECEMBER 1956

PAGE

A Model of Nuclear Shape Oscillations for γ-Transitions and Electron Excitation. By L. J. Tassie 407

Bounds for the Collapse Load of a Beam Compressed by Three Dies. By W. Freiberger 419

Calculation of Activation Energies from Calorimetric Work Carried Out at a Constant Rate of Heating. By J. F. Nicholas .. 425

Neutron Experiments with a Sensitive Szilard-Chalmers Detector. By R. D. Edge 429

The Behaviour of a Chapman Layer in the Night F_2 Region of the Ionosphere, under the Influence of Gravity, Diffusion, and Attachment. By R. A. Duncan 436

Motion in the Night-time E_s Region at Brisbane. By J. A. Thomas and M. J. Burke 440

High-frequency Backscatter Observations at Salisbury, South Australia. By C. G. McCue 454

Radio Echo Observations of Meteor Activity in the Southern Hemisphere. By C. D. Ellyett and C. S. L. Keay 471

A Survey of Cosmic Radio Emission at 600 Mc/s. By J. H. Piddington and G. H. Trent 481

An Investigation of the Radio Source 06N2A in Gemini. By H. Rishbeth 494

The Diurnal Variation of the Barometer Coefficient for Cosmic Rays at Hobart. By R. M. Jacklyn 505

Wind Profiles over the Sea and the Drag at the Sea Surface. By E. L. Deacon, P. A. Sheppard, and E. K. Webb 511

Statistical Analysis of Australian Temperature Data. By S. C. Das 542

A Relation between Snow Cover, Cirrus Cloud, and Freezing Nuclei in the Atmosphere. By E. G. Bowen 545

January Freezing Nucleus Measurements. By E. G. Bowen 552

Counts of Atmospheric Freezing Nuclei at Pretoria, January 1956. By S. C. Mossop, A. E. Carte, and K. J. Heffernan 556

Counts of Atmospheric Freezing Nuclei at Carnarvon, Western Australia, January 1956. By E. K. Bigg 561

Counts of Atmospheric Freezing Nuclei at Palo Alto, California, January 1956. By R. N. Bracewell 566

Short Communications

The Electronic Heat Capacity of a Copper-3-per cent. Zinc Alloy. By J. A. Rayne and W. R. G. Kemp 569

An Application of the Diffusion Equation to Viscous Motion with a Free Surface. By J. R. Philip 570

Double-layer Phenomena in the E_s Region. By J. A. Thomas .. 574

Index to Volume 9 579

LINEAR SUPERPOSITION IN VISCO-ELASTICITY AND THEORIES OF DELAYED EFFECTS

By E. R. LOVE*

[*Manuscript received August* 16, 1955]

Summary

Writers on elastic after-working or "visco-elasticity" have given a formula expressing the stress-history as an integral involving the strain-history linearly. If this is to be based directly on experiment, linear superposition is a necessary physical datum; but it is not sufficient, and examples are given to demonstrate this. Necessary and sufficient conditions are established in this paper, and expressed in physical as well as in mathematical terms; they may admit physical verification in some degree. Histories both with and without beginning are considered.

I. INTRODUCTION

Theories of hereditary phenomena generally use a relation between stimulus and response of the form

$$s(t)=\mu\sigma(t)+\int_0^t \varphi(t-\tau)\sigma(\tau)\mathrm{d}\tau, \qquad \ldots\ldots\ldots\ldots \quad \mathbf{(A)}$$

where $\sigma(t)$ measures the stimulus at time t, $s(t)$ the response at time t, μ is an instantaneous modulus, and $\varphi(\omega)$ is the intensity of memory over a time ω. Such a formula was propounded by Boltzmann (1876) for elastic after-working of metals, and further studied by Becker (1925); it was also developed by Volterra in studying hereditary phenomena in general. Some writers have preferred to use, in place of (**A**), the formula

$$s(t)=\int_0^t \sigma(t-\omega)\varphi(\omega)\mathrm{d}\omega. \qquad \ldots\ldots\ldots\ldots\ldots \quad \mathbf{(B)}$$

This is equivalent to (**A**) if $\varphi(\omega)$ is permitted to contain a δ-function.

In the context of visco-elasticity, (**A**) or (**B**) is usually based on assumptions regarding the micro-structure of the material (Gross 1947; Sips 1950; and others). However, theories of micro-structure are designed to explain the observations, and a less circuitous course would be to infer (**A**) directly from the observations. This has been suggested by I. M. Stuart in unpublished work to which he has kindly permitted me to refer.

If (**A**) holds, it is evident that linear superposition gives the response when two sets of stimuli operate together. But, if linear superposition is observed to hold, can we conversely infer (**A**) or (**B**)? Examples below show that this converse is false. The question thus arises: *What conditions, besides linear*

---

* Mathematics Department, University of Melbourne.

superposition, are necessary and sufficient to ensure that the stress-history s(t) *and the strain-history* σ(t) *are related by a formula such as* (**A**) *or* (**B**)*?*

This paper gives two answers to this question, one for strain-histories with a beginning and the other for those without. Most of the components of these answers are to be found in mathematical literature, and I am attempting to do little more than to assemble them.

II. LINEAR SUPERPOSITION

Linear superposition is not enough to ensure (**A**) or (**B**). With the meaning of "linear superposition" specified below, at L, the truth of this contention is demonstrated by the following examples.

(i) Suppose the material is viscous, with viscosity η, so that the strain-history $\sigma(t)$ is differentiable for all t and

$$s(t)=\eta\sigma'(t). \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{1})$$

Then the sum of two separate strain-histories clearly gives rise to the sum of the corresponding stress-histories, so that linear superposition holds. But there is no integrable function $\varphi(\omega)$ such that (**B**) holds. For, if there were, the strain-histories

$$\sigma_n(t)=0 \ \ (t\leqslant 0), \qquad \sigma_n(t)=\frac{\sin^2 nt}{n} \quad (t\geqslant 0)$$

would determine, for each positive integer n, the stress-history

$$s_n(t)=\int_0^t \varphi(\omega)\frac{\sin^2 n(t-\omega)}{n}d\omega \qquad (t\geqslant 0),$$

so that

$$|\, s_n(t)\,| \leqslant \frac{1}{n}\int_0^t |\,\varphi(\omega)\,|\, d\omega.$$

Thus $s_n(t)$ would tend to zero as $n\to\infty$, for each fixed value of t; whereas, on the contrary, (**1**) gives

$$s_n(t)=\eta \sin 2nt \qquad (t\geqslant 0).$$

If it is objected that (**1**) is inappropriate to this subject, as it makes stress independent of the previous history of strain, this objection cannot be raised against

$$s(t)=\eta\sigma'(t)+\varepsilon\int_0^t \sigma(\tau)d\tau,$$

but remarks similar to those above apply to this stress-strain relation also.

(ii) Suppose the stress-strain relation is

$$s(t)=\int_0^t \sigma(\tau)\tau d\tau. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{2})$$

Linear superposition clearly holds, but no relation (**B**). For if both held we should have, putting $\omega=t-\tau$ in (**B**),

$$\int_0^t \sigma(\tau)\{\varphi(t-\tau)-\tau\}\mathrm{d}\tau=s(t)-s(t)=0\ ;$$

and this would hold for every strain-history, so that

$$\varphi(t-\tau)-\tau=0 \qquad (0\leqslant\tau\leqslant t),$$

that is,

$$\varphi(\omega)=t-\omega \qquad (\omega\geqslant 0,\ t\geqslant 0).$$

Thus $\varphi(\omega)$ would have to be a function of t as well as of ω.

Similar remarks apply to the stress-strain relation

$$s(t)=\frac{2S}{t^2}\int_0^t \sigma(\tau)\tau\mathrm{d}\tau,$$

a plausible enough relation, if physical intuition can be trusted, since it makes the stress a weighted mean of the previous strains with the greatest weight attached to the most recent strains.

(iii) Suppose the stress-strain relation is expressed by (**4**) below, with $\psi(\omega)$ a continuous monotonic function which is not absolutely continuous (Riesz and Sz.-Nagy 1953, Section 24). Linear superposition holds for all piecewise-continuous strain-histories, but neither (**B**) nor (**A**) is capable of expressing the stress-strain relation. For if $\sigma(t)$ is the unit function defined in (**3**) below, the corresponding stress-history is $\psi(t)$ which is not absolutely continuous, but the right members of (**A**) and (**B**) are necessarily absolutely continuous.

On this account the use of a Stieltjes integral is inevitable for completeness, although in calculations which use only elementary functions it reduces to an ordinary integral. The Stieltjes integral also has the further advantage that no δ-function is necessary.

Additional conditions. In addition to linear superposition L, three other conditions, P, Q, and R, concerning stress and strain-histories are required. These are as follows.

L, *linear superposition: If two strain-histories* $\sigma_1(\mathrm{t})$ *and* $\sigma_2(\mathrm{t})$ *separately entail stress-histories* $\mathrm{s}_1(\mathrm{t})$ *and* $\mathrm{s}_2(\mathrm{t})$, *then the strain-history* $\sigma_1(\mathrm{t})+\sigma_2(\mathrm{t})$ *entails the stress-history* $\mathrm{s}_1(\mathrm{t})+\mathrm{s}_2(\mathrm{t})$.

Further, a history of no strain entails one of no stress.

P, *approximation: If strain-histories approximate one another until a certain instant, then so do the corresponding stress-histories until that instant.* More exactly:

If strain-histories $\sigma_n(t)$ approximate a strain-history $\sigma(t)$ in the sense of uniform convergence throughout $t\leqslant T$, that is,

$$\overline{\mathrm{bnd}}_{t\leqslant T}\ |\ \sigma_n(t)-\sigma(t)\ |\to 0 \qquad \text{as } n\to\infty,$$

where the upper bound is taken for all values of $t \leqslant T$, keeping n fixed; then, for the corresponding stress-histories $s_n(t)$ and $s(t)$,

$$s_n(t) - s(t) \to 0 \qquad \text{as } n \to \infty,$$

for each $t \leqslant T$.

This excludes stress-strain relations such as (**1**). It could be expressed in the language of functional analysis: the stress at any instant is a continuous functional of the previous strain-history, the distance between strain-histories being specified as above. Such conditions are essential in a theory of this kind.

Q, invariability: If one strain-history is a replica of another except for a steady delay throughout, then the corresponding stress-histories are in the same relation. (Cf. Volterra 1931, p. 189.) More exactly:

If, for some particular time of delay δ, the relation:

$$\sigma_1(t) = \sigma_2(t-\delta) \qquad \text{for all } t,$$

connects two strain-histories, then, for the corresponding stress-histories

$$s_1(t) = s_2(t-\delta) \qquad \text{for all } t.$$

This excludes stress-strain relations such as (**2**). It is a condition which might fail if the range of deformation were extensive enough to bring on elastic fatigue.

R, relaxation behaviour: The relaxation function vanishes for $t \leqslant 0$, *is continuous-on-the-left at all times, and is bounded.** *The relaxation function is here defined as the stress-history entailed by a strain-history of the unit-function continuous-on-the-left:*

$$u(t) = 0 \quad (t \leqslant 0), \qquad u(t) = 1 \quad (t > 0). \qquad \dots\dots\dots\dots \quad (\mathbf{3})$$

Physically it would no doubt be impossible to verify continuity-on-the-left, and immaterial; but some such assumption is needed in the theory to ensure the existence of all integrals occurring, unless we abandon such convenient fictions as discontinuous strain-histories. Throughout the paper the discontinuities of strain-histories will be similarly restricted, for the same reason.

III. HISTORIES BEGINNING AT $t=0$

In order that the stress-history $s(t)$ *be expressible, for every left-continuous strain-history* $\sigma(t)$ *vanishing throughout* $t \leqslant 0$, *by*

$$s(t) = \int_0^t \sigma(t-\omega)\,d\psi(\omega), \qquad \dots\dots\dots\dots\dots\dots \quad (\mathbf{4})$$

in which ψ *is a function of bounded variation on every finite interval and*

$$\psi(t) = 0 \qquad \textit{for all } t \leqslant 0,$$

it is necessary and sufficient that conditions L, P, Q, R hold for all left-continuous strain-histories vanishing throughout $t \leqslant 0$.

The function $\psi(t)$ *is the relaxation function defined in R above.*

---

* Boundedness is only needed later, for strain-histories having no beginning.

Before proving this theorem, the following notes on the meanings of terms are necessary.

A *left-continuous function* is continuous everywhere except for an isolated set of simple discontinuities-on-the-right. It is easily seen that the sum of two left-continuous functions is left-continuous, and that a left-continuous function is bounded in any finite interval.

The integral in (**4**) is to be understood as a slight extension of the elementary Stieltjes integral (Widder 1941, Ch. I), as follows. If $f(\omega)$ is continuous except for possible simple discontinuities at $c_0, c_1, \ldots, c_{k-1}, c_k$, where

$$a=c_0<c_1<\ldots<c_{k-1}<c_k=b,$$

we define

$$\int_a^b f(\omega)\mathrm{d}g(\omega)=\sum_{i=1}^{k}\int_{c_{i-1}}^{c_i} f(\omega)\mathrm{d}g(\omega), \quad \ldots\ldots\ldots\ldots\ldots \quad (5)$$

provided all the integrals on the right exist as elementary Stieltjes integrals. This definition is consistent with the elementary definition whenever the latter applies to the left side of (**5**), but has the advantage that integrability of f with respect to g is not necessarily destroyed if both are discontinuous at the one point, unless they are discontinuous on the same side of it.

(*a*) *Proofs that L, P, Q, R are necessary*

We suppose there is a function ψ, of bounded variation on every finite interval and vanishing for all $t\leqslant 0$, such that (**4**) expresses the stress-history for every " permitted " strain-history, namely those which are left-continuous and vanish throughout $t\leqslant 0$. We seek to deduce R, and also L, P, Q for all permitted strain-histories to which they are applicable.

Necessity of R. By our present hypothesis, the relaxation function defined in R above must be the stress-history $s(t)$ given by

$$s(t)=\int_0^t u(t-\omega)\mathrm{d}\psi(\omega), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

and this integral must exist for each t. By (**3**), the integrand is continuous except for a simple discontinuity-on-the-left at $\omega=t$; so the definition of the integral reduces to the elementary definition. If $t>0$, since the integral exists $\psi(\omega)$ must be continuous-on-the-left at $\omega=t$; and then

$$s(t)=\psi(t)-\psi(0)=\psi(t).$$

If $t\leqslant 0$ it follows immediately from (**3**) and (**6**) that

$$s(t)=0.$$

Thus R is established and ψ is identified with the relaxation function.

Necessity of L. For any two permitted strain-histories $\sigma_1(t)$ and $\sigma_2(t)$, the sum of the corresponding stress-histories $s_1(t)$ and $s_2(t)$ is

$$s_1(t)+s_2(t)=\int_0^t \sigma_1(t-\omega)d\psi(\omega)+\int_0^t \sigma_2(t-\omega)d\psi(\omega)$$
$$=\int_0^t \{\sigma_1(t-\omega)+\sigma_2(t-\omega)\}d\psi(\omega), \quad \dots\dots\dots\dots (7)$$

which, by our present hypothesis, is the stress-history entailed by the strain-history $\sigma_1(t)+\sigma_2(t)$, since this is left-continuous and vanishes for $t\leqslant 0$ and is therefore one to which (4) applies.

Necessity of P. Let $T>0$ and $\sigma_1(t)$, $\sigma_2(t)$, . . ., $\sigma(t)$ be permitted strain-histories such that

$$b_n=\underset{t\leqslant T}{\overline{\text{bnd}}}\ |\ \sigma_n(t)-\sigma(t)\ |\to 0 \qquad \text{as } n\to\infty.$$

For the corresponding stress-histories, if $0\leqslant t\leqslant T$,

$$|\ s_n(t)-s(t)\ |=\left|\int_0^t \{\sigma_n(t-\omega)-\sigma(t-\omega)\}d\psi(\omega)\ \right|$$
$$\leqslant b_n V(\psi\ ;\ 0,\ t), \quad \dots\dots\dots\dots\dots\dots\dots\dots (8)$$

where $V(\psi\ ;\ 0, t)$ is the total variation of ψ on $(0, t)$. Since this is independent of n, and $b_n\to 0$, P follows for $T>0$.

P is immediate for $T\leqslant 0$ since, by (4) and R,

$$s_n(t)=0=s(t) \qquad \text{for all } t\leqslant 0.$$

Necessity of Q. Let $\sigma_2(t)$ be left-continuous and vanish for $t\leqslant 0$; then $\sigma_1(t)=\sigma_2(t-\delta)$ is another such strain-history, supposing that $\delta>0$. For the corresponding stress-histories, at any particular instant $t>\delta$,

$$s_2(t-\delta)=\int_0^{t-\delta} \sigma_2(t-\delta-\omega)d\psi(\omega)$$
$$=\int_0^t \sigma_2(t-\omega-\delta)d\psi(\omega)=\int_0^t \sigma_1(t-\omega)d\psi(\omega)=s_1(t)\ ;$$

while for $t\leqslant\delta$ these equations still hold, every term being zero.

(*b*) *Proof that L, P, Q, R together are sufficient*

As before, we refer to strain-histories which are left-continuous and vanish throughout $t\leqslant 0$ as "permitted" strain-histories. We need first three lemmas. Only references or indications of proofs need be given for these.

Lemma 1. *If L and P hold for all permitted strain-histories, λ is any real number, and a permitted strain-history* σ(t) *entails a stress-history* s(t), *then the strain-history* $\lambda\sigma$(t) *entails the stress-history* λs(t).

This is proved successively for λ zero or a positive integer, positive rational, positive irrational, and negative real. Only for irrational values of λ is it necessary to use P.

Lemma 2. *If L and P hold for all permitted strain-histories, and* T *is fixed and positive, then the ratio*

$$| s(T) | \Big/ \overline{\text{bnd}}_{t \leqslant T} | \sigma(t) |$$

is bounded for all permitted strain-histories σ.

Proof as in Riesz and Sz.-Nagy (1953, p. 149), with $T\sigma$ (denoting a functional of σ) replaced by $s(T)$ and $\| \sigma \|$ replaced by $\overline{\text{bnd}}_{t \leqslant T} | \sigma(t) |$.

Lemma 3. *If* T *is fixed and positive, any permitted strain-history* $\sigma(\text{t})$ *can be approximated uniformly throughout* $\text{t} \leqslant \text{T}$, *that is, in the sense of P, by permitted strain-histories* $\sigma_n(\text{t})$ *which are also step-functions.*

In outline the proof is as follows. Referring to (**3**), the function

$$\sigma^*(t) = \sigma(t) - \sum_{j=0}^{l} \{\sigma(t_j+0) - \sigma(t_j)\} u(t-t_j)$$

is continuous in $0 \leqslant t \leqslant T$, the discontinuities of $\sigma(t)$ being at the points $t=t_j$ and being cancelled by those of the step-function subtracted from it. Thus $\sigma^*(t)$ is the uniform limit, in $0 \leqslant t \leqslant T$, of left-continuous step-functions $\sigma_n{}^*(t)$. Then

$$\sigma_n(t) = \sigma_n{}^*(t) + \sum_{j=0}^{l} \{\sigma(t_j+0) - \sigma(t_j)\} u(t-t_j)$$

are left-continuous step-functions converging to $\sigma(t)$ uniformly in $0 \leqslant t \leqslant T$. Further, they may be defined to be zero throughout $t \leqslant 0$, and these strain-histories $\sigma_n(t)$ have the requisite properties.

We can now proceed to prove the sufficiency of our conditions. We suppose that L, P, Q hold for all permitted strain-histories, and that R holds. We seek to establish that the relaxation function ψ, defined in R, is of bounded variation on every finite interval, and that (**4**) expresses the stress-history for every permitted strain-history; the proof is set out in these two stages. The former stage is essential to the latter; the integral in (**4**) will not exist for all permitted strain-histories unless ψ is of bounded variation.

The relaxation function ψ *is of bounded variation on every finite interval.* The essentials of the following proof are given in Riesz and Sz.-Nagy, p. 109. This is part of the proof of Riesz's important theorem expressing linear functionals as Stieltjes integrals. It is presented below in a simpler form; simpler because the central difficulty in Riesz's theorem is the extension of a functional outside its domain of definition, whereas in our context this extension is already provided by the relaxation function ψ.

To show that ψ is of bounded variation on an interval $(0, T)$, consider any subdivision $0 = \omega_0 < \omega_1 < \ldots < \omega_k = T$, write

$$\tau_i = T - \omega_i \qquad (i = 0, 1, \ldots, k),$$

and construct a left-continuous strain-history vanishing for $t \leqslant 0$:

$$\sigma(t) = \lambda_i \text{ in } \tau_i < t \leqslant \tau_{i-1} \; (i = 1, 2, \ldots, k),$$

where

$$\lambda_i = \begin{cases} +1 \text{ if } \psi(\omega_i)-\psi(\omega_{i-1})>0, \\ 0 \text{ if } \psi(\omega_i)-\psi(\omega_{i-1})=0, \\ -1 \text{ if } \psi(\omega_i)-\psi(\omega_{i-1})<0. \end{cases} \qquad \text{............} \quad (\mathbf{9})$$

This function is expressible in terms of the unit function (**3**) :

$$\sigma(t) = \sum_{i=1}^{k} \lambda_i \{u(t-\tau_i)-u(t-\tau_{i-1})\}. \qquad \text{............} \quad (\mathbf{10})$$

By R, a strain-history $u(t)$ entails stress-history $\psi(t)$.

By Q, a strain-history $u(t-\delta)$ entails stress-history $\psi(t-\delta)$.

By lemma 1, a strain-history $\lambda u(t-\delta)$ entails the stress-history $\lambda\psi(t-\delta)$; and by L it follows that the strain-history (**10**) entails the stress-history

$$s(t) = \sum_{i=1}^{k} \lambda_i \{\psi(t-\tau_i)-\psi(t-\tau_{i-1})\}. \qquad \text{..................} \quad (\mathbf{11})$$

Then

$$s(T) = \sum_{i=1}^{k} \lambda_i \{\psi(\omega_i)-\psi(\omega_{i-1})\} = \sum_{i=1}^{k} |\, \psi(\omega_i)-\psi(\omega_{i-1}) \,|,$$

and

$$|\, s(T) \,| \Big/ \overline{\mathrm{bnd}}_{t\leqslant T} \,|\, \sigma(t) \,| = \sum_{i=1}^{k} |\, \psi(\omega_i)-\psi(\omega_{i-1}) \,|.$$

For fixed $T>0$, and any subdivision of $(0, T)$, this expression is subject to the bound established in lemma 2. Thus ψ is of bounded variation on $(0, T)$, and so on any finite interval.

The stress-history is given by (**4**) *for every permitted strain-history.* To prove this we choose a permitted strain-history $\sigma(t)$ and an instant $T>0$, and consider the approximating strain-histories $\sigma_n(t)$ provided by lemma 3.

Since each $\sigma_n(t)$ is a step-function, and left-continuous, it can be expressed in terms of the unit-function (**3**) in the form

$$\sigma_n(t) = \sum_{i=1}^{k_n} \lambda_{ni} \{u(t-t_{ni-1})-u(t-t_{ni})\},$$

where

$$0 = t_{n0} < t_{n1} < \ldots < t_{nk_n} = T$$

and λ_{ni} is the value of $\sigma_n(t)$ throughout $t_{ni-1}<t\leqslant t_{ni}$. The corresponding stress-history is

$$s_n(t) = \sum_{i=1}^{k_n} \lambda_{ni} \{\psi(t-t_{ni-1})-\psi(t-t_{ni})\},$$

by applications of R, Q, lemma 1, and L analogous to those made in arguing from (**10**) to (**11**).

At time $t=T$ the above stress can be written

$$s_n(T) = -\sum_{i=1}^{k_n} \int_{t_{ni-1}}^{t_{ni}} \sigma_n(t) \mathrm{d}\psi(T-t),$$

the Stieltjes integrals existing in the elementary sense because, in $t_{ni-1} \leqslant t \leqslant t_{ni}$, $\sigma_n(t)$ is constant except for discontinuity-on-the-right at $t = t_{ni-1}$ while $\psi(T-t)$ is continuous-on-the-right by R. Using the Stieltjes integral defined in (**5**), it follows that

$$s_n(T) = -\int_0^T \sigma_n(t) \mathrm{d}\psi(T-t). \quad \ldots\ldots\ldots\ldots (\mathbf{12})$$

It remains only to let $n \to \infty$ in (**12**). The left side then tends to $s(T)$, by P, where $s(t)$ denotes the stress-history entailed by the given strain-history $\sigma(t)$. The right side is expected to tend to

$$-\int_0^T \sigma(t) \mathrm{d}\psi(T-t), \quad \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{13})$$

but before proving this we must know that this integral exists. It does exist in the sense of (**5**), for the component integrals between adjacent discontinuities of the integrand exist in the elementary Stieltjes sense; this is because $\psi(T-t)$ is of bounded variation and has no discontinuity-on-the-right while $\sigma(t)$ is continuous except for a terminal discontinuity-on-the-right. Now

$$\left| \int_0^T \sigma(t) \mathrm{d}\psi(T-t) - \int_0^T \sigma_n(t) \mathrm{d}\psi(T-t) \right| \leqslant b_n V(\psi;\ 0,\ T),$$

where

$$b_n = \overline{\underset{t \leqslant T}{\mathrm{bnd}}} \mid \sigma(t) - \sigma_n(t) \mid;$$

and this tends to zero as $n \to \infty$, by lemma 3. Thus the right side of (**12**) indeed tends to (**13**), and, by P,

$$s(T) = -\int_0^T \sigma(t) \mathrm{d}\psi(T-t) = \int_0^T \sigma(T-\omega) \mathrm{d}\psi(\omega).$$

So (**4**) is established at any instant $t > 0$. To establish it for $t \leqslant 0$, it is clear that the integral is then zero, since ψ is constant in the range by R; so we have only to ensure that the stress-history $s(t)$ entailed by any permitted strain-history $\sigma(t)$ vanishes throughout $t \leqslant 0$. This is a consequence of P and R, obtained by taking, in the former, $T = 0$ and

$$\sigma_n(t) = u(t) \qquad \text{for each } n.$$

For then

$$\overline{\underset{t \leqslant 0}{\mathrm{bnd}}} \mid \sigma_n(t) - \sigma(t) \mid = \overline{\underset{t \leqslant 0}{\mathrm{bnd}}} \mid \sigma(t) \mid,$$

which vanishes because $\sigma(t)$ is a permitted strain-history; therefore, by P,

$$s_n(t) \to s(t) \text{ as } n \to \infty, \text{ throughout } t \leqslant 0.$$

But $s_n(t) = \psi(t)$, which vanishes throughout $t \leqslant 0$, by R; consequently $s(t)$ does the same.

IV. STRAIN-HISTORIES HAVING NO BEGINNING

The preceding theorem applies only to strain-histories vanishing throughout $t \leqslant 0$, and is not properly applicable to strain-histories having no beginning, such as those of steady oscillation. An extra condition is needed, akin to Volterra's principle of dissipation of hereditary action (Volterra 1931, p. 188) but not obviously equivalent to it.

S, decay: Suppression of sufficiently ancient parts of a strain-history has practically no effect on the present stress. More exactly:

If strain-histories $\sigma_n(t)$ have agreed with a given strain-history $\sigma(t)$ since $t=-n$, so that

$$\sigma_n(t)=\sigma(t) \quad (t>-n), \qquad \sigma_n(t)=0 \quad (t\leqslant -n),$$

then, for the corresponding stress-histories $s_n(t)$ and $s(t)$,

$$s_n(t)\to s(t) \quad \text{as } n\to\infty, \text{ for each fixed } t.$$

The following theorem, analogous to that of Section III, can now be proved.

In order that the stress-history s(t) *be expressible, for every bounded left-continuous strain-history* σ(t), *by*

$$s(t)=\int_0^\infty \sigma(t-\omega)d\psi(\omega), \qquad \dots\dots\dots\dots\dots\dots \quad \mathbf{(14)}$$

in which ψ *is a function of bounded variation on* (0, ∞), *it is necessary and sufficient that conditions L, P, Q, R, S hold for all bounded left-continuous strain-histories.*

The function ψ(t) *is still the relaxation function defined in R.*

The infinite integral in (**14**) is to be understood as

$$\lim_{T\to\infty}\int_0^T \sigma(t-\omega)d\psi(\omega),$$

where the integral on $(0, T)$ is understood in the sense of (**5**), the sense so far used in this paper.

Proofs that L, P, Q, R are necessary proceed almost as before, considering all bounded left-continuous strain-histories instead of only those previously permitted. We suppose that (**14**) expresses the stress-history corresponding to each such strain-history, the function ψ being of bounded variation on (0, ∞) and not merely on each finite interval. In the course of these proofs this function ψ is again identified with the relaxation function. The only other differences are that infinite integrals occur in (**6**), (**7**), and (**8**). That in (**6**) reduces at once, if $t>0$, to the finite integral already considered; those in (**7**) and (**8**) are treated as before, the latter giving

$$| s_n(t)-s(t) | \leqslant b_n V(\psi;\ 0,\ \infty),\ \to 0 \text{ as } n\to\infty,$$

for each $t\leqslant T$ whether T is positive or not.

Proof that S necessarily holds for bounded left-continuous strain-histories. If $\sigma(t)$ and $\sigma_n(t)$ are such strain-histories, related as in the hypothesis of S, (**14**) gives

$$\begin{aligned} | s(t)-s_n(t) | &= \left| \int_{t+n}^{\infty} \sigma(t-\omega)d\psi(\omega) \right| \\ &\leqslant KV\ (\psi;\ t+n,\ \infty), \end{aligned}$$

where K is a bound of $|\sigma(t)|$. For any fixed t this expression tends to zero as $n \to \infty$, so that S holds.

Proof that L, P, Q, R, S together are sufficient, supposing that they hold for all bounded left-continuous strain-histories. These conditions hold, in particular, for all bounded left-continuous strain-histories vanishing throughout $t \leqslant 0$; so the stress-strain relation (**4**) applies in all such cases, by the preceding theorem in a form slightly modified so as to consider only bounded strain-histories.

Now suppose $\sigma(t)$ is any bounded left-continuous strain-history. Let $\sigma_n(t)$ be the strain-histories related with it as in S, and let $s(t)$ and $s_n(t)$ be the corresponding stress-histories. Since $\sigma_n(t-n)$ is a bounded left-continuous strain-history which vanishes for $t \leqslant 0$, and $s_n(t-n)$ is the corresponding stress-history by Q, it follows from (**4**) that

$$s_n(t-n) = \int_0^t \sigma_n(t-\omega-n)d\psi(\omega),$$

for all t; whence, replacing t by $t+n$,

$$s_n(t) = \int_0^{t+n} \sigma_n(t-\omega)d\psi(\omega).$$

Considering this equation with t fixed and $n > -t$, the integrand can be replaced by $\sigma(t-\omega)$, since this function agrees with it throughout the range except possibly at $\omega = t+n$, and at this terminal value $\psi(\omega)$ is continuous-on-the-left by R. Thus

$$s_n(t) = \int_0^{t+n} \sigma(t-\omega)d\psi(\omega), \quad \ldots\ldots\ldots\ldots\ (\mathbf{15})$$

for each t and each positive integer $n > -t$.

From (**15**) we now prove that ψ is of bounded variation on $(0, \infty)$. If it were not so, there would be a succession of instants ω_i, starting with $\omega_0 = 0$, such that

$$\sum_{i=1}^{m} | \psi(\omega_i) - \psi(\omega_{i-1}) | \to \infty \text{ as } m \to \infty. \quad \ldots\ldots\ (\mathbf{16})$$

Using these instants ω_i we construct a bounded left-continuous strain-history

$$\sigma(t) = \lambda_i \text{ in } -\omega_i < t \leqslant -\omega_{i-1} \qquad (i=1,\ 2, \ldots),$$

in which λ_i is 0 or ± 1 exactly as in (**9**); then, by (**15**),

$$\begin{aligned} s_n(0) &= \int_0^n \sigma(-\omega)d\psi(\omega) \\ &= \sum_{i=1}^{k} | \psi(\omega_i) - \psi(\omega_{i-1}) | + \sigma(-n)\{\psi(n) - \psi(\omega_k)\}, \end{aligned}$$

where k is the greatest integer i such that $\omega_i < n$; properly k should be denoted k_n, and $k_n \to \infty$ as $n \to \infty$. But the last term on the right is bounded, by R; so that, by (**16**),

$$s_n(0) \to \infty \qquad \text{as } n \to \infty.$$

This contradicts S; and so we must admit that ψ is of bounded variation on $(0, \infty)$.

We can now easily deduce that (**14**) holds for every bounded left-continuous strain-history $\sigma(t)$. In particular, the integral in (**14**) exists in the sense stated; for, if K is a bound of $|\sigma(t)|$,

$$\left| \int_T^{T'} \sigma(t-\omega)\mathrm{d}\psi(\omega) \right| \leqslant KV(\psi;\ T,\ T') \leqslant KV(\psi;\ T,\ \infty)$$

supposing $T' > T$; and this tends to zero as $T \to \infty$. Thus (**15**) gives, for each fixed t,

$$s_n(t) \to \int_0^\infty \sigma(t-\omega)\mathrm{d}\psi(\omega) \qquad \text{as } n \to \infty;$$

from which (**14**) follows by S.

V. References

Becker, R. (1925).—Elastische Nachwirkung und Plastizität. *Z. Phys.* **33**: 185.

Boltzmann, L. (1876).—Zur Theorie der elastischen Nachwirkung. *Ann. Phys., Lpz.* **7**: 624.

Gross, B. (1947).—On creep and relaxation. *J. Appl. Phys.* **18**: 212.

Riesz, F., and Sz.-Nagy (1953).—"Leçons d'analyse fonctionnelle." (Acad. Sci. Hongrie: Budapest.)

Sips, R. (1950).—Mechanical behaviour of visco-elastic substances. *J. Polymer Sci.* **5**: 69.

Volterra, V. (1931).—"Theory of Functionals." Translation by M. Long. pp. 147, 188, 192, 217. (Blackie: London.)

Widder, D. V. (1941).—"The Laplace Transform." (Princeton Univ. Press.)

RECIPROCAL STATIC SOLUTIONS OF THE EQUATIONS OF THE GRAVITATIONAL FIELD

By H. A. Buchdahl*

[*Manuscript received September* 23, 1955]

Summary

This paper deals with reciprocal static line-elements, previously defined by the author, the condition that their Ricci tensors vanish being no longer imposed. If the indices i, k run from 1 to n with the exception of the fixed index a (the line-elements being static with respect to x^a) a certain quantity Q^ν_μ appears with the remarkable property that in a reciprocal transformation Q^k_i is invariant, whilst Q^a_a merely changes sign. Q^ν_μ is closely related to the Hamiltonian derivative of the Gaussian curvature, so that the general results obtained may be applied to the field equations of General Relativity Theory, with $n=a=4$. Q^4_4 is then the total energy density; and formally every static distribution of matter has a " reciprocal distribution " associated with it. In particular, the equation of state of a distribution of fluid reciprocal to a distribution of fluid possessing a given equation of state may be obtained directly from the latter, i.e. without the solution of the field equations being known.

I. Introduction

In a previous paper (Buchdahl 1954), hereafter referred to as RI, the line-element *reciprocal* to the static line-element

$$ds^2=g_{ik}(x^j)dx^i dx^k+g_{aa}(x^j)(dx^a)^2 \quad \ldots\ldots\ldots\ldots \quad (1.1)$$

was defined to be

$$ds^2=(g_{aa})^{2/(n-3)}g_{ik}dx^i dx^k+(g_{aa})^{-1}(dx^a)^2. \quad \ldots\ldots \quad (1.2)$$

Here roman indices run over the range 1, . . ., n ($\geqslant 4$) with the omission of the *fixed* index a. It was shown that, if (**1.1**) satisfies the equations

$$G_{\mu\nu}=0, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.3)$$

then so does (**1.2**), $G_{\mu\nu}$ being the Ricci tensor of the V_n the metric of which is (**1.1**). (Greek indices run over the entire range 1, . . ., n.) In particular, when $n=4$ and a is chosen to be 4, (**1.3**) are the equations of the gravitational field in empty space.

I now consider reciprocal static line-elements without assuming them to satisfy (**1.3**). In particular it will be shown (Section II) that, if

$$P^\nu_\mu=\tfrac{1}{2}\delta^\nu_\mu G-G^\nu_\mu, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1.4)$$

---

* Physics Department, University of Tasmania.

and P^{ν}_{μ} is the corresponding tensor-density, then the identities

$$\left.\begin{aligned}&\bar{\mathrm{P}}^k_i=\mathrm{P}^k_i,\\&\left(\bar{\mathrm{P}}^a_a-\frac{1}{n-3}\bar{\mathrm{P}}^i_i\right)=-\left(\mathrm{P}^a_a-\frac{1}{n-3}\mathrm{P}^i_i\right)\end{aligned}\right\}\quad\ldots\ldots\ldots\ (\mathbf{1.5})$$

hold, unbarred and barred quantities being associated with (**1.1**) and (**1.2**) respectively. When $n=a=4$ it then follows from (**1.5**) that if one has any formal* static solution of the gravitational field equations in the presence of matter, then one can immediately write down another solution which is "reciprocal" to it (Section III (a)), and the total energies of two such distributions, supposed finite and regular, are equal and opposite in sign (Section III (b)). In particular, if one contemplates "reciprocal distributions"† of fluids in thermodynamic equilibrium (Section IV) whose equations of state are $\rho=f(p)$ and $\bar{\rho}=\bar{f}(\bar{p})$ respectively, then the second of these equations may be deduced from the first without either of the corresponding solutions of the field equations being known (Section V), and two such equations of state will be said to be reciprocal to one another.

II. The Ricci Tensors of Reciprocal Metrics

It was shown in RI, Section 2, that the components of the Ricci tensor belonging to the static line-element

$$\mathrm{d}s^2=\mathrm{e}^{2\sigma}\underset{\circ}{g}_{ik}\mathrm{d}x^i\mathrm{d}x^k+\mathrm{e}^{2\gamma}(\mathrm{d}x^a)^2\quad\ldots\ldots\ldots\ldots\ (\mathbf{2.1})$$

may be written

$$\left.\begin{aligned}&G_{ik}=\underset{\circ}{G}_{ik}+(n-3)(\sigma_{;ik}-\sigma_{;i}\sigma_{;k})+\gamma_{;ik}+\gamma_{;i}\gamma_{;k}-(\sigma_{;i}\gamma_{;k}+\sigma_{;k}\gamma_{;i})\\&\qquad\qquad+\underset{\circ}{g}_{ik}[\Delta\sigma+(n-3)\widehat{\sigma\sigma}+\widehat{\sigma\gamma}],\\&G_{ia}=0,\\&G_{aa}=\mathrm{e}^{2\gamma-2\sigma}[\Delta\gamma+\widehat{\gamma\gamma}+(n-3)\widehat{\sigma\gamma}].\end{aligned}\right\}\ldots\ (\mathbf{2.2})$$

Here a subscript following a semicolon denotes covariant differentiation in the $\underset{\circ}{V}_{n-1}$ whose metric is $\mathrm{d}s^2=\underset{\circ}{g}_{ik}\mathrm{d}x^i\mathrm{d}x^k$, and $\underset{\circ}{G}_{ik}$ is the Ricci tensor of this V_{n-1}, whilst if ξ, η are any two scalars, $\widehat{\xi\eta}$ stands for $\underset{\circ}{g}^{ik}\xi_{;i}\eta_{;k}$ and $\Delta\xi$ for $\underset{\circ}{g}^{ik}\xi_{;ik}$. From (**1.4**) and (**2.2**) one then has

$$\left.\begin{aligned}&P_{ik}=\underset{\circ}{P}_{ik}-(n-3)(\sigma_{;ik}-\sigma_{;i}\sigma_{;k})-(\gamma_{;ik}+\gamma_{;i}\gamma_{;k})+(\sigma_{;i}\gamma_{;k}+\sigma_{;k}\gamma_{;i})\\&\qquad+g_{ik}[(n-3)\Delta\sigma+\Delta\gamma+\tfrac{1}{2}(n-3)(n-4)\widehat{\sigma\sigma}+(n-4)\widehat{\sigma\gamma}+\widehat{\gamma\gamma}],\\&P_{ia}=0,\\&P_{aa}=\mathrm{e}^{2\gamma-2\sigma}\left(\frac{1}{n-3}\underset{\circ}{P}+(n-2)\Delta\sigma+\tfrac{1}{2}(n-2)(n-3)\widehat{\sigma\sigma}\right).\end{aligned}\right\}\ldots\ (\mathbf{2.3})$$

---

* See the remarks at the beginning of Section III (b).

† I.e. distributions such that the respective solutions $g_{\mu\nu}$, $\bar{g}_{\mu\nu}$ of the field equations are reciprocal to one another.

The tensor $P_{\mu\nu}$ associated with the V_n whose metric is (**1.1**) is obtained from (**2.3**) by setting $\sigma=0$, $\gamma=\tau=\frac{1}{2}\log g_{aa}$ and omitting the subscript 0 :

$$\left.\begin{aligned} P_{ik}&=P'_{ik}-(\tau_{;ik}+\tau_{;i}\tau_{;k})+g_{ik}(\Delta\tau+\widehat{\tau\tau}),\\ P_{ia}&=0,\\ P_{aa}&=\frac{1}{n-3}e^{2\tau}P'. \end{aligned}\right\} \tag{2.4}$$

Here P'_{ik} is associated with the V_{n-1} whose metric is $ds^2=g_{ik}dx^idx^k$. Similarly, setting $\sigma=2\tau/(n-3)$ and $\gamma=-\tau$ in (**2.3**) one has the tensor $\bar{P}_{\mu\nu}$ associated with $\bar{V}_n$, that is, the V_n whose line-element is reciprocal to (**1.1**). Thus

$$\left.\begin{aligned} \bar{P}_{ik}&=P'_{ik}-(\tau_{;ik}+\tau_{;i}\tau_{;k})+g_{ik}(\Delta\tau+\widehat{\tau\tau}),\\ \bar{P}_{ia}&=0,\\ \bar{P}_{aa}&=\frac{1}{n-3}\exp\left(-\frac{2(n-1)}{n-3}\tau\right)[P'+2(n-2)(\Delta\tau+\widehat{\tau\tau})]. \end{aligned}\right\} \tag{2.5}$$

Raising one of the indices in each of the equations (**2.4**) and (**2.5**) it then follows at once that

$$\left.\begin{aligned} \bar{P}^k_i&=(g_{aa})^{-2/(n-3)}P^k_i,\\ \bar{P}^a_a&=-(g_{aa})^{-2/(n-3)}\left(P^a_a-\frac{2}{n-3}P^i_i\right). \end{aligned}\right\} \tag{2.6}$$

The components P^ν_μ, $\bar{P}^\nu_\mu$ in which just one index equals a are of course zero, and they will not be explicitly referred to hereafter. Keeping in mind that

$$\sqrt{(-\bar{g})}=(g_{aa})^{2/(n-3)}\sqrt{(-g)}, \tag{2.7}$$

the equations (**1.5**) follow at once. If one defines

$$Q^\nu_\mu=P^\nu_\mu-\frac{1}{n-3}\delta^\nu_\mu P^i_i, \tag{2.8}$$

then (**2.6**) may be written in the curious form

$$\bar{Q}^k_i=Q^k_i,\quad \bar{Q}^a_a=-Q^a_a. \tag{2.9}$$

In transformations of the x^i, Q^k_i behaves as a tensor and Q^a_a as a scalar.

III. RECIPROCAL DISTRIBUTIONS OF MATTER

(*a*) If one chooses $n=4$, the preceding results may be applied to the field equations of General Relativity Theory which are (in suitable units)

$$P^\nu_\mu=T^\nu_\mu, \tag{3.1}$$

T^ν_μ being the components of the stress-energy tensor of the medium filling the region of interest. x^a will be taken to be the coordinate time x^4. The T^ν_μ

and $\bar{T}^{\nu}_{\mu}$ associated with any metric and its reciprocal are then related as follows, according to (**2.6**):

$$\bar{T}^{k}_{i}=(g_{44})^{-2}T^{k}_{i},\quad \bar{T}^{4}_{4}=-(g_{44})^{-2}(T^{4}_{4}-2T^{i}_{i}). \quad \ldots\ldots\ldots\ldots \quad (\mathbf{3.2})$$

The components of the stress-energy tensor $\underset{\circ}{T}_{\mu\nu}$ as measured by an observer at rest with respect to the coordinates employed above, who uses proper coordinates $\underset{\circ}{x}^{\mu}$ at any point of interest, are (Tolman 1934, p. 215)

$$\underset{\circ}{T}_{ik}=p_{ik},\quad \underset{\circ}{T}_{i4}=0,\quad \underset{\circ}{T}_{44}=\rho, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{3.3})$$

ρ, p_{ik} being the density and mechanical stresses of the medium under consideration. The results of measurements* carried out by such an observer on distributions represented by a metric tensor $g_{\mu\nu}$ on the one hand, and by its reciprocal $\bar{g}_{\mu\nu}$ on the other are, in view of (**3.2**), therefore related by

$$\bar{p}_{ik}=(g_{44})^{-2}p_{ik},\quad \bar{\rho}=-(g_{44})^{-2}(\rho-2\sum_{i=1}^{3}p_{ii}). \quad \ldots\ldots \quad (\mathbf{3.4})$$

(*b*) From (**3.4**), or more directly from (**2.9**), it follows that of two finite, regular reciprocal distributions T^{ν}_{μ} and $\bar{T}^{\nu}_{\mu}$ only one can be physically realizable. A result due to Tolman (1934, p. 234) states that the total energy U of such a distribution of matter, including the energy of the field produced by it, is given by

$$U=\int \mathrm{Q}^{4}_{4}\mathrm{d}x_{1}\mathrm{d}x_{2}\mathrm{d}x_{3}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{3.5})$$

If $\bar{U}$ is the total energy of the "reciprocal distribution", it follows at once from the second of equations (**2.9**) that

$$U=-\bar{U}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{3.6})$$

The stated result is implied whenever it is granted that any physically realizable distribution must have positive total energy.

IV. FLUID DISTRIBUTIONS

An interesting special case is that of a distribution of fluid in thermodynamic equilibrium. For this

$$T^{k}_{i}=-p\delta^{k}_{i},\quad T^{4}_{i}=T^{i}_{4}=0,\quad T^{4}_{4}=\rho, \quad \ldots\ldots\ldots\ldots \quad (\mathbf{4.1})$$

where ρ and p are the proper density and pressure of the fluid. Equations (**3.2**) then read

$$\bar{p}=(g_{44})^{-2}p,\quad \bar{\rho}=-(g_{44})^{-2}(\rho+6p). \quad \ldots\ldots\ldots \quad (\mathbf{4.2})$$

---

* See the remarks at the beginning of Section III (*b*).

Now the equation

$$T^{\nu}_{\mu;\nu}=0, \qquad (4.3)$$

which is identically satisfied, can here be reduced to the form (Tolman 1934, p. 317)

$$\text{grad } p+\tfrac{1}{2}(\rho+p) \text{ grad} \log g_{44}=0. \qquad (4.4)$$

Let the equation of state of the fluid be given in the form $\rho=f(p)$. Assuming $\rho+p\neq 0$, (**4.4**) then gives at once

$$\log g_{44}=-2\int dp/(\rho+p). \qquad (4.5)$$

Equations (**4.2**) may therefore be written

$$\bar{p}=p \exp \int\frac{4dp}{\rho+p}, \quad \bar{\rho}=-(\rho+6p) \exp \int\frac{4dp}{\rho+p}. \qquad (4.6)$$

These equations are a parametric representation of the equation of state* of the fluid constituting the reciprocal distribution.

V. Reciprocal Equations of State

One may look upon (**4.6**) as directly defining an equation of state $\bar{\rho}=\bar{f}(\bar{p})$ which is " reciprocal to " the given equation of state $\rho=f(p)$. Note that the explicit form of the distribution, *qua* function of the coordinates, is not required. A particularly simple example is furnished by the equation of state

$$\rho=sp, \qquad (5.1)$$

where $s(\neq-1)$ is a constant. Then

$$\bar{p}=\beta p^{c}, \quad \bar{\rho}=-(s+6)\beta p^{c}, \qquad (5.2)$$

where $c=(s+5)/(s+1)$, and β is a positive constant. Hence

$$\bar{\rho}=-(s+6)\bar{p}. \qquad (5.3)$$

The equation $\bar{\rho}=-3\bar{p}$ is evidently reciprocal to itself.

VI. Electrostatic Fields

If T^{ν}_{μ} is the stress-energy tensor of an electrostatic field, then the same cannot be true of $\bar{T}^{\nu}_{\mu}$. For otherwise the vanishing of the spur of the electromagnetic stress-energy tensor would require

$$T^{i}_{i}=-T^{4}_{4} \text{ and } \bar{T}^{i}_{i}=-\bar{T}^{4}_{4} \qquad (6.1)$$

to hold simultaneously. By (**3.2**) the first of these implies

$$\bar{T}^{4}_{4}=3\bar{T}^{i}_{i}, \qquad (6.2)$$

which is inconsistent with (**6.1**) since T^{4}_{4} cannot vanish.

---

* Actually they define a one-parameter set of equations of state since a generally non-trivial constant of integration occurs in (**4.6**). A change of this constant is, however, simply equivalent to a change of units in which $\bar{\rho}$ and $\bar{p}$ are measured.

VII. Inclusion of the Cosmical Constant

The possible presence of a cosmological term in the field equations has hitherto been ignored. If this term be restored, (**3.1**) must be replaced by

$$P_{\mu}^{\nu}-\delta_{\mu}^{\nu}\lambda=T_{\mu}^{\nu}, \qquad \qquad (\mathbf{7.1})$$

where λ is the cosmical constant. Equations (**3.2**) are therefore appropriately modified by replacing T_{μ}^{ν} by $T_{\mu}^{\nu}+\delta_{\mu}^{\nu}\lambda$ throughout. In the same way, equations (**4.6**) become

$$\bar{p}=(p-\lambda)J+\lambda, \quad \bar{\rho}=-(\rho+6p-5\lambda)J-\lambda, \qquad \qquad (\mathbf{7.2})$$

where

$$J=\exp 4\int dp/(\rho+p). \qquad \qquad (\mathbf{7.3})$$

VIII. References

Buchdahl, H. A. (1954).—*Quart. J. Math.* **5**: 116.

Tolman, R. C. (1934).—" Relativity, Thermodynamics and Cosmology." (Oxford Univ. Press.)

THE ACCELERATED MOTION OF DROPLETS AND BUBBLES

By T. Pearcey* and G. W. Hill*

[*Manuscript received October* 6, 1955]

Summary

The drag force acting upon small particles such as droplets and bubbles, moving through a viscous medium, depends upon the rate of change of the state of motion of the medium and upon the diffusion of vorticity from the surface of the particle. For accelerated particles the drag force changes with time and depends upon all previous accelerations. The equation of motion of a particle takes the form of an integro-differential equation, which has been solved numerically in the case of deceleration to rest from uniform motion with no impressed body forces. In any experiments designed according to the principles of dynamical similarity, the ratios of viscosities and densities of the medium and the particle must be maintained constant in the scale transformations involved.

I. Introduction

Recent studies of cloud and droplet physics have raised the question of what may be the detailed behaviour of very small droplets subject to forces which change with time ; in particular whether or not it is possible to simulate the motion of such droplets in the laboratory by suitable changes in the scale of the variables (Sartor 1954). In theoretical studies of droplet dynamics it has so far been usual to assume that the only drag force experienced by a droplet, moving through a viscous medium at any instant with a Reynolds number R, to be identical with the drag it would experience if it were moving uniformly at that same Reynolds number, and the acceleration has been assumed to depend only upon the difference between this drag force and any time-variable body forces acting upon the droplet only.

The actual accelerations suffered by a body, immersed in a viscous medium and subject to time-variable forces, are affected by two factors. As is well known, the momentum imparted by such forces is distributed between the body and the surrounding medium, giving rise to an effect which, in a non-viscous medium, is equivalent to an increase in the mass of the body. In fact, for a spherical droplet the effective increase in mass is equal to half the mass of the medium displaced by the droplet (Lamb 1932, p. 124). A further effect occurs owing to the finite rate at which vorticity diffuses from the surface of the body. The distribution of vorticity throughout the medium depends upon the past velocity of the body and thus upon its history. The actual drag experienced at any particular time is more affected by the recent past history than by the distant past.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

The case of the motion of a lamina under given forces and immersed in a viscous medium has been discussed previously by others (Lamb 1932, p. 591). The variation with time of the effect of past motions upon the current drag force for a sphere is here found to be similar to the case of a lamina.

In the case of droplets and bubbles the effect of past history on the equations of rectilinear motion is determined to a degree of approximation corresponding to that given by Oseen's approximation to the problem of viscous flow around a sphere. The resulting integro-differential equation shows that any experiment, which endeavours to simulate the case of droplets and bubbles by changing the scales of the physical constants involved, must maintain constant both the Reynolds number and the ratios of viscosities and densities of the droplet and medium.

The equation of motion has been solved numerically in the case of a droplet, initially moving with uniform velocity, which is allowed to come to rest as a result of the drag forces arising from the surrounding medium.

The solutions show that the effect of past history on the motion of a small droplet or bubble may in certain circumstances affect, by a relatively large amount, its velocity and the distance it travels so that neglect of this effect in theoretical studies would cause serious error.

II. The Drag Force for Impulsive Motion of a Droplet

We shall restrict consideration at first to the case of a droplet of radius a supposed to be at rest for $t<0$ and to move with uniform velocity U for $t>0$. The drag force experienced by the droplet will be the same as that which would apply if the droplet were held at rest and the medium allowed to flow past it with a main stream velocity $-U$.

The flow may be expressed in terms of a Stokes type of stream function, Ψ, and Oseen's approximation to the inertia terms in the hydrodynamical equations for small U may be adopted. Thus Ψ satisfies the partial differential equation

$$\left.\begin{aligned}&\left(\nu D^2-\frac{\partial}{\partial t}-U\frac{\partial}{\partial x}\right)D^2\Psi=0,\\ \text{where}\qquad &D^2\equiv\frac{\partial^2}{\partial r^2}+\frac{\sin\theta}{r^2}\;\frac{\partial}{\partial\theta}\left(\frac{1}{\sin\theta}\;\frac{\partial}{\partial\theta}\right),\end{aligned}\right\}\quad\ldots\ldots\ (1)$$

ν is the kinematic viscosity of the medium, and r,θ,φ are spherical polar coordinates with the centre of the droplet as origin. The velocity U is directed along the positive axis of x corresponding to $\theta=0$. For $t>0$ we may assume the component of the stream function corresponding to the uniform motion of the medium, namely $-Ur^2\sin^2\theta/2$, to be omitted, and restrict attention to the disturbance caused by the droplet.

The components of fluid velocity caused by the presence of the droplet are

$$u_r=\frac{1}{r^2\sin\theta}\;\frac{\partial\Psi}{\partial\theta},\quad u_\theta=-\frac{1}{r\sin\theta}\;\frac{\partial\Psi}{\partial r},$$

and the boundary conditions, assuming no slipping to occur at the surface for $t > 0$, become

$$\begin{aligned} u_r &= U\cos\theta, \; u_\theta = -U\sin\theta, && r = a, \\ u_r &= u_\theta = 0, && r = \infty. \end{aligned}$$

The remaining components of the stream function may be expressed as the sum of an irrotational function $\Psi^{(1)}$ and a solenoidal function $\Psi^{(2)}$, each of which is obtainable in the form of an infinite series of functions which satisfy the boundary conditions at infinity and whose coefficients are determined by the conditions at the surface of the droplet.

The irrotational component is of impulsive type, whilst the solenoidal component corresponds to the diffusion of vorticity from the surface of the droplet required to maintain the boundary conditions. Both components provide contributions to the drag experienced by the droplet which behave differently with time.

Taking Laplace transforms in the variable p, $\overline{\Psi^{(1)}}$ and $\overline{\Psi^{(2)}}$, where $\overline{\Psi} = \overline{\Psi^{(1)}} + \overline{\Psi^{(2)}}$, then

$$\left.\begin{aligned} \overline{\Psi^{(1)}} &= \sum_{n=1}^{\infty} A_n(p)/r^n . (1-\mu^2)P'_n(\mu), \\ \overline{\Psi^{(2)}} &= e^{\mu k r} \sum_{n=1}^{\infty} B_n(p)\chi_n(qr)(1-\mu^2)P_n(\mu), \end{aligned}\right\} \quad \ldots\ldots \; (2)$$

where

$$\begin{aligned} \chi_n(z) &= z^{\frac{1}{2}} K_{n+\frac{1}{2}}(z), \\ q^2 &= k^2 + p/\nu, \\ k &= U/2\nu = R/4a. \end{aligned}$$

R is the Reynolds number $2Ua/\nu$, K_m is the modified Bessel's function of second kind of order m, and $P_n(\mu)$ is the Legendre function of order n.

The boundary conditions require that

$$A_n + a^n \sum_{m=1}^{\infty} B_m \chi_m(qa) X_{nm}(ka) = U a^{n+2} \delta_{1n}/2p,$$

$$\sum_{m=1}^{\infty} B_m [n\chi_m(qa)X_{nm}(ka) + ka\chi_m(qa)X_{nm}'(ka) + qa\chi_m'(qa) + X_{nm}(ka)]$$
$$= 3Ua^2\delta_{1n}/2p, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \; (3)$$

where the δ_{1n} is the Kronecker function and X_{nm} is the function introduced by Goldstein (1929), and is defined by the identity

$$e^{\mu z} P_m'(\mu) = \sum_{n=1}^{\infty} X_{nm}(z) P_n'(\mu). \quad \ldots\ldots\ldots\ldots\ldots\ldots \; (4)$$

Solution of the equations for the A_n and B_n, followed by inversion of the Laplace transform, would provide the solution for the flow pattern of the stream function. So far as the drag on the sphere is concerned we require only to integrate the appropriate stress components over the surface of the droplet.

Thus we integrate the axial component of the tangential or viscous stress $p_{r\theta}$ and the normal or pressure stress p_{rr}, where

$$\left.\begin{aligned} p_{rr} &= -p+2\eta\frac{\partial u_r}{\partial r}, \\ p_{r\theta} &= \eta\left\{r\frac{\partial}{\partial r}\left(\frac{u_\theta}{r}\right)+\frac{1}{r}\frac{\partial u_r}{\partial \theta}\right\}, \end{aligned}\right\} \quad r=a,$$

p is the hydrostatic pressure, and η is the coefficient of viscosity of the medium. The total drag force D is then given by

$$D=2\pi a^2\int_0^\pi(-p_{rr}\cos\theta+p_{r\theta}\sin\theta)\sin\theta\, d\theta.$$

Expressing this in terms of the stream function at $r=a$, and using the relation between the hydrostatic pressure p and the stream function,

$$\frac{\partial p}{\partial \theta}=\frac{\rho}{\sin\theta}\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial x}\right)\frac{\partial \Psi^{(1)}}{\partial r},$$

where ρ is the density of the medium and, taking Laplace transforms, then

$$\bar{D}=\frac{4\pi\rho a}{3}\left[Ua^2-\frac{3p}{a}A_1\right]+\frac{8\pi\rho U}{5a^2}A_2. \quad \dots\dots\dots\dots (5)$$

Actually a complete determination of D requires knowledge of all the A's and B's in order that A_1 and A_2 may be obtained. In the case of small droplets some approximations may be made. Thus A_1 will be of order U, and A_2 of order U^2. For sufficiently small U, A_2 may be neglected and in the determination of A_1 all terms for $n>1$ may be omitted. Thus we find

$$A_1=(Ua^3/2p)[1-3/\{1+qa\chi'(qa)/\chi(qa)+kaX_{11}'(ka)/X_{11}(ka)\}],$$

where

$$\chi_1(z)=\surd(\tfrac{1}{2}\pi)e^{-z}(1+z^{-1}),$$
$$X_{11}(z)=z^{-\frac{1}{2}}I_{3/2}(z),$$

and $I_{3/2}(z)$ is the modified Bessel function of first kind and order 3/2. For sufficiently small ka, therefore,

$$kaX_{11}'/X_{11}=1+k^2a^2/5+O(k^4a^4).$$

If, however, k^2a^2 be neglected throughout compared with $(p/\nu)a$, then upon carrying out the inversion on the Laplace transformation, the drag force reduces to

$$D=6\pi\eta aU[1+a/\surd(\pi\nu t)]+Um\delta(t)/2, \quad t\geqslant 0, \quad \dots\dots (6)$$

where m is the mass of the medium displaced by the droplet and $\delta(t)$ is the Dirac δ-function. Thus D, the force required instantaneously to raise a droplet from rest to uniform motion, consists of three terms: (1) the viscous force which would be experienced in the case of uniform translation, in this case the drag according to Stokes' theory of a uniformly translated droplet; (2) a time-variable component which is infinite at $t=0$ and decreases as $t^{-\frac{1}{2}}$; and (3) a transient force which occurs only at $t=0$ and is due to the momentum transferred to the medium by the sudden change in velocity.

It will be shown later, in Section IV, that the equation of motion of a droplet, depending upon (**6**) for the drag, implies physically inadmissible conditions. Although a droplet, subject to no body forces after time $t=0$, may come to rest, (**6**) implies that it must travel an infinite distance before doing so, owing to the complete neglect of terms of the order k^2a^2. If k^2a^2 be retained, the behaviour differs for large elapsed times and the distances travelled become finite. Thus we find

$$D=6\pi\eta a U\left[1+\frac{1}{\sqrt{(\pi T)}}-\frac{1}{\sqrt{\pi}}\,\frac{(1+\gamma_1)\gamma_1{}^2}{(\gamma_1-\gamma_2)}\exp\left(-\gamma_1^2T\right)\int_0^T\exp\left(\gamma_1^2T\right)\mathrm{d}T/\sqrt{T}\right.$$
$$\left.+\frac{1}{\sqrt{\pi}}\,\frac{(1+\gamma_2)\gamma_2^2}{(\gamma_1-\gamma_2)}\exp\left(-\gamma_2^2T\right)\int_0^T\exp\gamma_2^2T\,\mathrm{d}T/\sqrt{T}\right]+\tfrac{1}{2}Um\delta(T),\quad \ldots(\mathbf{7})$$

where $T=\nu t/a^2$ and γ_1 and γ_2 are the roots of the quadratic

$$80z^2-R^2z+4R^2=0.$$

For small T, D approximates to

$$D=6\pi\eta a U\left[1+\frac{1}{\sqrt{(\pi T)}}\left\{1+\frac{3R^2}{40}T+\mathrm{O}(R^4T^2)\right\}\right]+\tfrac{1}{2}Um\delta(T),\qquad 0\leqslant T<16/R^2,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\quad(\mathbf{8.1})$$

which corresponds to (**6**), and for large T becomes

$$D=6\pi\eta a U[1+(\pi T)^{-\frac{1}{2}}\{25/2R^2T+\mathrm{O}(R^{-4}T^{-2})\}],\qquad T>16/R^2,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\quad(\mathbf{8.2})$$

thus, for $T>16/R^2$, D decreases more rapidly than predicted by (**6**).

III. THE IMPULSIVE MOTION OF BUBBLES

The main assumption made in Section II was that of rigidity of the droplet. This applies only if the droplet is highly viscous compared with the surrounding medium, and does not apply to bubbles. When internal motions are considered, the stream function in the interior is represented by Ψ_1, and its Laplace transform is

$$\overline{\Psi}_1=\sum_1^\infty[C_nr^{n+1}+\mathrm{e}^{-k\mu r}E_n\psi_n(q_1r)](1-\mu^2)P_n'(\mu),\qquad 0\leqslant r\leqslant a,$$

in which the unit suffix indicates reference to the variables and physical constants of the medium of the bubble, and ψ_n involves the modified Bessel function of first kind

$$\psi_n(z)=z^{\frac{1}{2}}I_{n+\frac{1}{2}}(z).$$

The boundary conditions at the surface require the viscous stresses and the velocities of the two media to be continuous, that is,

$$u_r(a)=0,$$

$$\operatorname*{Lim}_{\varepsilon=0}[u_\theta(a+\varepsilon)-u_\theta(a+\varepsilon)]=0,$$

$$\operatorname*{Lim}_{\varepsilon=0}\left[\eta\frac{\partial}{\partial a}u_r(a+\varepsilon)-\eta_1\frac{\partial}{\partial a}u_r(a-\varepsilon)\right]=0,$$

and

$$\operatorname*{Lim}_{\varepsilon=0}\left[\eta\left\{\frac{\partial}{\partial a}-\frac{1}{a}\right\}u_\theta(a+\varepsilon)-\eta_1\left\{\frac{\partial}{\partial a}-\frac{1}{a}\right\}u_\theta(a-\varepsilon)\right]=0.$$

These conditions are sufficient to determine the coefficients A_n, B_n, C_n, and E_n, although complete evaluation would be impracticable.

If the drag force be evaluated to the same degree of approximation as was adopted in equation (**6**) we find

$$D=6\pi\eta aU\xi[1+a/\sqrt{(\pi\nu t)}]+Um\delta(t)/2,\quad t\geqslant 0, \quad \ldots\ldots\ldots \quad (\mathbf{9})$$

where ξ is a parameter depending only upon the coefficients of viscosity of the two media, that is,

$$\xi=(3\eta_1+2\eta)/(3\eta_1+3\eta).$$

The form of the expression (**9**) is the same as that for a rigid droplet, although in this case the component due to viscosity is modified by the factor ξ which varies between the limit 1, which it takes in the case of highly viscous droplets ($\eta_1\gg\eta$), and 2/3 in the case of bubbles ($\eta\gg\eta_1$). This factor arises from the internal motions which allow flow to occur at the surface, thus reducing the viscous stresses. The drag force in the case of uniform translation (when $t=\infty$) agrees with that quoted by Lamb (1932, pp. 600–1). The force arising from the transfer of momentum to the surrounding fluid remains unchanged to the degree of approximation adopted. This is as might be expected since the bubble is presumed to retain its shape during the impulsive change of velocity. This would be so in practice if the internal pressure due to surface tension were large enough.

IV. THE EQUATIONS OF MOTION OF DROPLETS AND BUBBLES

To the approximation adopted, the drag force is a linear function of the velocity U, and consequently, for continuously variable velocities, the method of superposition of solutions may be adopted in which U is replaced by the increment in velocity over an infinitesimally small interval of time between T and $T+\mathrm{d}T$, i.e. $\dot{U}(t-T)\mathrm{d}T$, the summation being taken over all previous elapsed time for which U is non-zero. Making the substitution for the drag force we find the equation of rectilinear motion to be

$$(M+\tfrac{1}{2}m)\dot{U}+6\pi\eta a\xi U+\frac{6\pi\eta a^2\xi}{\sqrt{(\pi\nu)}}\int_0^\infty\frac{\dot{U}(t-t')}{\sqrt{t'}}\mathrm{d}t'=F,$$

where F represents the applied or body force acting upon the mass, M, of the droplet or bubble. The term $\frac{1}{2}m$ constitutes the effective increase in mass arising from the transfer of momentum to the medium as the acceleration occurs and is not due to viscosity. The equation of motion may be reduced to non-dimensional form by the transformations

$$\left.\begin{aligned} &\tau=\nu\pi\lambda^2 t/\xi^2 a^2, \quad \lambda^2=9\rho\xi/\pi(2\rho_1+\rho), \\ \text{and}\quad &V=u/U_0, \qquad\quad f=F/6\pi\eta a\xi U_0, \end{aligned}\right\} \quad \ldots\ldots\ldots \; \mathbf{(10.1)}$$

where U_0 is some standard velocity, whence

$$\frac{\mathrm{d}V}{\mathrm{d}\tau}+V+\lambda\int_0^\infty \frac{\mathrm{d}V(\tau-\tau')}{\mathrm{d}(\tau-\tau')}\cdot\frac{\mathrm{d}\tau'}{\sqrt{\tau'}}=f. \quad \ldots\ldots\ldots \; \mathbf{(10.2)}$$

It is clear that the motion of a small droplet or bubble at any instant will depend markedly upon its previous motions, since the effect upon the drag force of past accelerations decreases slowly with elapsed time t, as $t^{-\frac{1}{2}}$, only so long as the maximum Reynolds number is sufficiently small (cf. (**8.1**) and (**8.2**)). The actual magnitude of the effect depends upon the relative densities of the two media and their coefficients of viscosity, and any experiment conducted and designed according to principles of dynamical similarity must be arranged so that the value of the parameter λ is maintained constant during the transformation of scale.

The value of λ is affected only little by variations of the viscosity of the medium since $\xi^{\frac{1}{2}}$ may vary only between the limits $(\frac{2}{3})^{\frac{1}{2}}$ and 1. In this respect the need to maintain a constant ratio of viscosity may well be relaxed. However, the variation with the density is important as λ may be about $(9\rho/2\pi\rho_1)^{\frac{1}{2}}$ for droplets ($\rho_1 \gg \rho$), and $(6/\pi)^{\frac{1}{2}}$ for bubbles. In particular,

$$\begin{aligned} \lambda&=0\cdot 0428, &&\text{for water droplets in air,} \\ &=1\cdot 38, &&\text{for air bubbles in water.} \end{aligned}$$

V. Solution of the Equations of Motion

To illustrate the effect of past history upon the velocity and position of droplets and bubbles, a simple situation is considered in which, up to the time $t=0$, motion is uniform and rectilinear and determined by a constant impressed body force such as that due to gravity. At time $t=0$ and for all subsequent time this force is supposed removed, so that the particle decelerates to rest under the action of viscous forces.

Suppose therefore that U_0 be taken so that f is equal to unity for $\tau<0$ and zero thereafter. In the absence of any effects of past acceleration, i.e. $\lambda=0$, the velocity for $\tau>0$ would be $\mathrm{e}^{-\tau}$. The difference in the velocity due to the effect of past history, ω, satisfies the integro-differential equation

$$\frac{\mathrm{d}\omega}{\mathrm{d}\tau}+\omega+\lambda\int_0^\tau \frac{\mathrm{d}\omega(\tau-\tau')}{\mathrm{d}(\tau-\tau')}\,\frac{\mathrm{d}\tau'}{\sqrt{\tau'}}=\lambda\mathrm{e}^{-\tau}\int_0^\tau \mathrm{e}^{\tau'}\frac{\mathrm{d}\tau'}{\sqrt{\tau'}}, \quad \tau>0, \quad \ldots \; \mathbf{(11.1)}$$

and

$$\omega=0, \text{ at } \tau=0.$$

The difference in the distance travelled will be linearly related to s, where

$$s=\int_0^{\tau} \omega d\tau. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{11.2})$$

Solution of the equation (**11**) for ω as a power series in τ gives an expression which is useful when τ is sufficiently small, thus

$$\omega=\sum_{n=0}^{\infty} a_n \tau^{2n+3/2},$$

where

$$a_0=\frac{4\lambda}{3}, \qquad a_1=-X,$$

$$a_2=\frac{8\lambda}{15}(1-X), \qquad a_3=X\left(1-\frac{2X}{3}\right),$$

$$a_4=\frac{16\lambda}{35}\left(1-\frac{8X}{3}+\frac{4X^2}{3}\right), \qquad a_5=-\frac{X}{2}\left(1-\frac{5X}{3}+\frac{7X^2}{6}\right),$$

and $X=\frac{1}{2}\pi\lambda^2$. Expansion into power series of λ gives

$$\omega=\sum_{r=1}^{\infty} \omega_r \lambda^r,$$

where

$$\omega_r=-e^{-\tau}\int_0^{\tau} e^{\mu}\int_0^{\mu} \omega'_{r-1}(\mu-\tau')\frac{d\tau'}{\sqrt{\tau'}} \cdot d\mu, \qquad r>1,$$

and

$$\omega_1=e^{-\tau}\int_0^{\tau}\int_0^{\mu} e^{\tau'}\frac{d\tau'}{\sqrt{\tau'}}d\mu.$$

For large elapsed times, $\tau\gg1$, ω_1 is at most, of the order $\tau^{-\frac{1}{2}}$, and from the differential equation it is seen that the ω_r is then of order $\tau^{-(r+1)/2}$, at most. Thus

$$\omega=\lambda\tau^{-\frac{1}{2}}+O(\lambda^2\tau^{-1},\ \tau^{-3/2}), \qquad \tau\gg1.$$

The solution for ω has been carried out numerically for various values of λ. The computations were assisted by transforming the variables as follows

$$\tau=\sigma^2, \qquad \tau'=\sigma'^2,$$

and adjusting the degree of the singularity of the integrand by writing

$$\omega'(\sigma)=-\frac{2\sigma}{(1+\pi\lambda\sigma)}\left\{\omega(\sigma)+\lambda\left[\int_0^{\sigma}\frac{\{\omega'(\sigma')-\omega'(\sigma)\}}{\sqrt{(\sigma^2-\sigma'^2)}} \cdot d\sigma'-2F(\sigma)\right]\right\},$$

where

$$F(\sigma)=\exp(-\sigma^2)\int_0^{\sigma}\exp(\sigma'^2)d\sigma',$$

and $\omega'(\sigma)=d\omega/d\tau$, etc.

The numerical integration was carried out on the C.S.I.R.O. digital computer using a process similar to that due to Huen, and using three stages of approximation for each interval h of σ ($h=1/50$). As the values of ω and ω' were found at the end of each interval they were recorded and ω' stored for use in the computation of all succeeding values. The computation of each value proceeded in three stages. Thus, $\omega'^{(1)}_N$ was found by extrapolation

$$\omega'^{(1)}_N=2\omega'_{N-1}-\omega'_{N-2},$$

whence by integration,

$$\omega^{(1)}_N=\omega_{N-1}+\tfrac{1}{2}h(\omega'_{N-1}+\omega'^{(1)}_N),$$

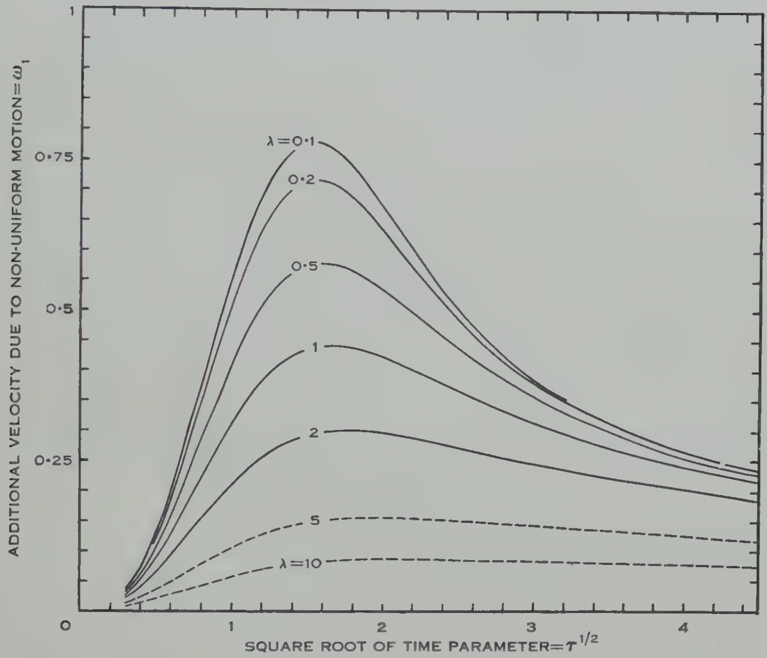


Fig. 1.—The effect upon the velocity of a decelerated droplet or bubble of past non-uniform motion. The curves are solutions of equation (**11.1**).

from which $\omega'^{(2)}_N$ was found by use of the differential equation where the integral was computed to the same degree of precision as in the integration, thus

$$\omega'^{(2)}_N=-\frac{2Nh}{(1+\pi\lambda Nh)}\left[\omega^{(1)}_N+\lambda\left\{\sum_{n=0}^{N-1}\frac{(1+\delta_{0N})}{2}\ \frac{(\omega'_N-\omega'^{(1)}_N)}{\sqrt{(N^2-n^2)}}-2F_N\right\}\right],$$

in which F_N, taken for values at the interval h, were supplied from punched tape, and δ_{0N} is the Kronecker δ-function. Using $\omega'^{(2)}_N$ a new value of ω_N was found by integration, from which the final value of ω'_N was obtained

$$\omega'_N=\omega'^{(3)}_N=-\frac{2Nh}{(1+\lambda\pi Nh)}\cdot\left[\overline{\omega^{(2)}_N-\omega^{(1)}_N}+\tfrac{1}{2}\pi\lambda\overline{\omega'^{(1)}_N-\omega'^{(2)}_N}\right].$$

A final integration over the interval gave ω_N, truncation errors being of order h^{-3} only. The distance travelled, s_N, at the end of the Nth interval was obtained by integration following the evaluation of ω_N, thus

$$s_N = s_{N-1} + (\tfrac{1}{2}h)^2(2N-1)(\omega_{N-1}+\omega_N).$$

Figure 1 shows the ratio ω/λ, for values of λ exceeding the range of the physically important cases. The curves are plotted against the abscissa σ which is proportional to the square root of the elapsed time.

TABLE 1

THE MAXIMUM SHOWN EFFECT OF PAST HISTORY UPON THE VELOCITY $\omega_{\text{max.}}$ OF A DROPLET OR BUBBLE AT ELAPSED TIME τ

λ	$\omega_{\text{max.}}$	τ
0·1	0·0780	1·233
0·2	0·1433	1·240
0·5	0·2893	1·259
1·0	0·4406	1·287
2·0	0·6011	1·327
5·0	0·7705	1·389
10·0	0·8692	1·439

Table 1 gives the values of the maximum change in velocity due to the history of the motions. In particular it is seen that the difference may be as much as 0·5 for the maximum physically admissible value of λ. This occurs at an elapsed time $\tau = 1 \cdot 30$.

VI. The Distance Travelled

In the absence of effects due to history, the distance travelled would be unity. The excess distance travelled, s, is of order $\tau^{5/2}$ for small τ, but for large τ, according to equation (**6**) of order $\tau^{\frac{1}{2}}$. Such a behaviour for large elapsed times does not occur in practice. The actual distance travelled before coming to rest is finite. The theoretical result is due to the neglect of terms of order R^2 in the calculation of the drag.

If allowance be made for terms of this order the integrand of the equation of motion for large τ' would, from equation (**8.2**), become of order $\tau'^{-3/2}R^{-2}$ instead of $\tau'^{-\frac{1}{2}}$. Figure 2 shows the excess distance travelled as the ratio s/λ, on the basis of equation (**6**) for the drag force. The steadily increasing behaviour of s for large elapsed times is evident.

To allow for the inclusion of the drag force for large elapsed times, the integrand of equation (**11**) must be multiplied by a factor $25\pi\lambda^2/2R^2\tau$. Thus ω would decrease as $\tau^{-3/2}$ and s would vary as $\tau^{-\frac{1}{2}}$ and tend to a constant limit. However, the solution obtained here may be considered as a reasonable approximation for $\tau \leqslant 16/R^2$, where R is the Reynolds number of the droplet or bubble at its initial velocity.

A rough estimate of the total distance travelled may be obtained by assuming the motion to be controlled by the weighting factor $\tau^{-\frac{1}{2}}$ up to the time at which the additional factor entering into the integrand is of the order unity, at which time the motion ceases. This time $\tau_{\text{max.}}$ may be taken as

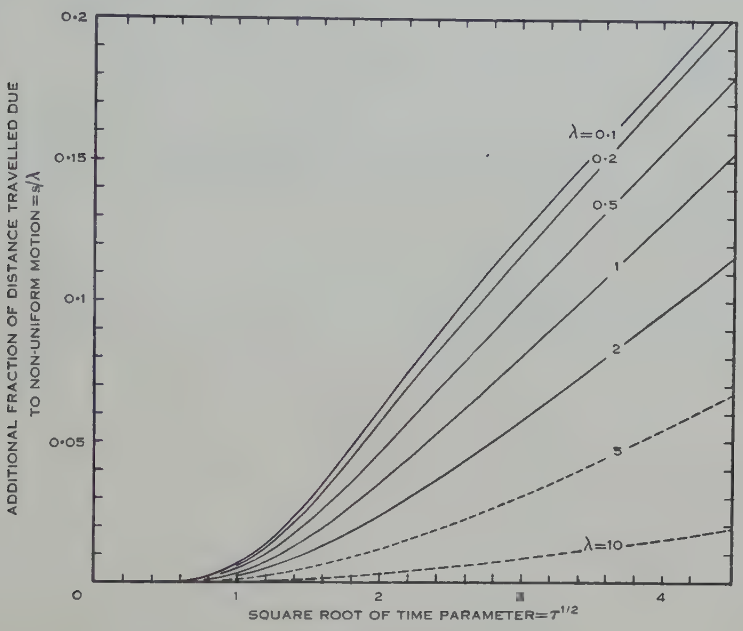


Fig. 2.—The additional distance travelled by a decelerated droplet or bubble due to the effect of non-uniform motion. The curves are solutions of equations (**11.1**) and (**11.2**).

$$\tau_{\text{max.}} \simeq 36\lambda^2/R^2.$$

TABLE 2

APPROXIMATE MAXIMUM EXCESS DISTANCE $s_{\text{max.}}$ TRAVELLED FOR A DROPLET OR BUBBLE

λ	$s_{\text{max.}}/\lambda$
0·1	$-0\cdot0411+0\cdot0332/R$*
0·2	$-0\cdot0533+0\cdot0683/R$
0·5	$-0\cdot0597+0\cdot160/R$
1·0	$-0\cdot0631+0\cdot287/R$
2·0	$-0\cdot0540+0\cdot447/R$

* R=Reynolds number.

Table 2 gives the approximate values of the maximum excess distance assuming the curves of Figure 2 to apply up to $\tau_{\text{max.}}$.

VII. Conclusion

The effect of the finite rate of the diffusion of vorticity in the case of accelerated motion of small droplets and bubbles can be an important factor in determining at any instant their velocity and position. Although the effect is small for liquid droplets immersed within a gas, it is important in the case of vapour bubbles in a liquid or for two immiscible liquids.

Any attempt to carry out laboratory experiments to simulate the case of small droplets or bubbles must preserve a strict dynamical similitude which extends to maintaining the correct value of the parameter λ under the conditions of the experiment.

VIII. References

Goldstein, S. (1929).—*Proc. Roy. Soc.* A **123** : 225–35.
Lamb, H. (1932).—" Hydrodynamics." (Cambridge Univ. Press.)
Sartor, D. (1954).—*J. Met.* **11** : 91–103.

GROWING ELECTRIC SPACE-CHARGE WAVES

By J. H. PIDDINGTON*

[*Manuscript received September* 5, 1955]

Summary

A theory of growing space-charge electric waves, due to Pierce, Haeff, and others, is thought to explain the operation of Haeff's electron-wave tube and other amplifying devices and perhaps the origin of some solar radio emission. The theory is shown to be untenable, the growth predicted being spurious and due to misinterpretation of the dispersion equations.

Two rules are given which should be observed when interpreting dispersion equations :

(*a*) The frame of reference in which the dispersion equation is developed should be stationary in the gas in which the waves are propagated.

(*b*) When choosing real or imaginary parts of frequency or propagation constant for the dispersion equation, the choice must be consistent with physically realizable and relevant conditions.

An alternative theory of operation of the electron-wave and other growing-wave tubes is given and some design factors are briefly discussed.

I. INTRODUCTION

The results described in this paper may be applicable to any type of wave but refer specifically to longitudinal electron oscillations in a neutral ion-electron plasma. These are called space-charge, electron pressure, or plasma oscillations. They were first investigated as standing waves (Tonks and Langmuir 1929), but a more general theory of travelling space-charge waves is now available (Thomson and Thomson 1933 ; Bohm and Gross 1949*a*).

A theory of spontaneous " growth " of space-charge waves in interpenetrating electron streams or mixed ion and electron streams has been developed by Haeff (1948, 1949*a*, 1949*b*), Nergaard (1948), Pierce (1948, 1949, 1950), Pierce and Herbenstreit (1948), Bohm and Gross (1949*b*), Feinstein and Sen (1951), Rydbeck and Forsgren (1951), and others. The theory purports to show how the waves " grow " or steadily increase in amplitude as they propagate along the composite electron stream. It is believed to explain the operation of Haeff's electron-wave tube and (in varied forms) other growing-wave amplifying devices and perhaps also the origin of some solar radio emission.

It is the purpose of the present paper to show that the theory is not tenable, the growth predicted being spurious and due to misinterpretation of the dispersion equations. It is possible that the same criticism may apply to the theory of operation of the travelling-wave tube (Pierce 1950) although here the position

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

is complicated by the presence of a metal helix in the electron stream. A similar criticism applies (Piddington 1955) to a theory of growing transverse electromagnetic (radio) waves advanced by Bailey (1948, 1950, 1951, and other references). These errors indicate the desirability of some rules to use in the interpretation of dispersion equations when dealing with waves in moving gas ; two rules are suggested.

Since the electron-wave tube demonstrably *does* work, it behoves a critic of the theory to suggest an alternative mechanism. Possible modes of growth of space-charge waves are briefly discussed and a theory of the electron-wave tube, and perhaps other amplifying devices, is suggested.

II. THE SUBSTITUTION ANALYSIS

The theory of the electron-wave tube depends on a substitution analysis in which certain types of solution are sought of the relevant equations : Maxwell's electromagnetic field relationships and the equations of motions of electrons in two or more interpenetrating streams having different drift velocities. The solutions have the form $\exp \mathrm{i}\,(\omega t-kx)$ and describe plane waves moving in time t and space x. In general both the frequency ω and propagation constant k may be complex quantities : $\omega=\omega_r+\mathrm{i}\omega_i$, $k=k_r+\mathrm{i}k_i$, where ω_r, ω_i, k_r, k_i are real. The wave is then given by $\exp\,(k_ix-\omega_it)\,.\,\exp\,\mathrm{i}(\omega_rt-k_rx)$ and so may grow or decay in both time and space. The result of the substitution analysis is a dispersion equation relating ω and k but not giving either directly. The derivation of this equation is, apart from occasional mathematical complexities, generally not difficult. It is in its interpretation that trouble has been met.

By equating real and imaginary parts, the dispersion equation provides two relationships between the four quantities concerned. Hence two of the quantities must be chosen more or less arbitrarily. The procedure from this point has been somewhat a matter of physical intuition guided by mathematical checks. In many cases, notably the magneto-ionic theory of radio waves due to Appleton, Hartree, and others, the results have proved eminently successful. However, when dealing with an electron gas which moves relative to the observer some serious errors have been made and definite rules of procedure are desirable.

In the magneto-ionic theory it is usual to assume $\omega_i=0$ and assign some particular value to ω_r. This means, physically, that the emitter, operating at a given frequency, is steady and stationary relative to the observer ; otherwise the observed wave amplitude would change with time ($\omega_i=0$). Within one frequency range the value of k is then found to be complex with real and imaginary parts of opposite sign, indicating a wave which decays in space. The decay is associated with a scattering of electrons by collisions with heavy ions, atoms, or molecules, thus destroying their ordered motion. Within another frequency range k is imaginary, indicating stationary or reflected waves. These are described more fully below.

A similar procedure has been followed in the electron-wave tube theory but here values of k are found within certain wave bands which indicate wave *growth*. This is interpreted as indicating an increase in wave energy at the expense of the kinetic energy of the electron streams. Pierce (1950) has called it an electromechanical process. A somewhat similar theory is used to explain the operation of the travelling-wave tube (Pierce 1950).

Twiss (1951) has sensed a danger in this interpretation of the dispersion equation. For a double electron stream he has shown that if k (rather than ω) is assumed real, then the equation leads to a complex value of ω, which might be interpreted as a wave growing in time. Such growth is not observed experimentally, so that doubt is cast on the theory. He concludes that a theory of growing waves may only be developed in relation to the boundaries of the medium which are essential factors in promoting growth. It is shown below that neither Twiss's criticism of the theory nor his emphasis of the role of boundaries is justified.

However, there is a fundamental error in the theory of the electron-wave tube, due to misinterpretation of the dispersion equation. When correctly interpreted no real wave growth is found under any conditions. The main reason for the error is that the wave equations are derived in a system of axes moving relative to the electron gas.

III. THE MOVING OBSERVER

Space-charge electric waves comprise perturbations of an electron gas together with an (electric) potential field due to the perturbations. They are propagated *relative to the gas itself* and, if this drifts relative to the observer, it carries the wave with it so that the latter assumes different apparent properties. It is easy to see that a steady, spatially attenuated wave might, in a suitable reference system, appear to grow in time. It is less obvious, but will now be shown, that other waves may appear to grow in space when no real growth is present.

To illustrate spurious wave growth, the simplest and best-known example of space-charge waves is chosen : waves in a gas whose electrons have thermal random motions but no mass drift. The dispersion equation (Thomson and Thomson 1933 ; Bohm and Gross 1949*a*) is

$$v_t^2k^2=\omega^2-\omega_0^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where v_t is of the order of the electron thermal velocity and ω_0 is the resonance frequency of the plasma, and collisions of electrons with heavy ions or atoms have been neglected. When $\omega > \omega_0$ loss-free travelling waves propagate with velocity $v_t\omega(\omega^2-\omega_0^2)^{-\frac{1}{2}}$. When $\omega < \omega_0$, k is imaginary and a pair of waves of the form $\exp k_ix \,.\, \exp i\omega_rt$ occur near the emitter. These are standing, exponentially spatially attenuated or " evanescent " waves and are, in effect, waves which are rejected by the medium and are being reflected back into the emitter. It is with these waves that we are concerned ; a pair of them is shown schematically in Figure 1 (*a*).

At no frequency do any of these waves grow as they propagate. The only spatial intensity change is due to a process of reflection.

Now consider these same waves as seen by an observer moving relative to the gas with velocity U along the x axis. The waves are now described by the constants ω_1, k_1 given by the Lorentz transformation :

$$\omega=\beta(\omega_1+Uk_1) \quad \text{and} \quad k=\beta\left(k_1+\frac{U}{c^2}\omega_1\right), \qquad (2)$$

where $\beta=(1-U^2/c^2)^{-\frac{1}{2}}$. When $U \ll c$ and the wave velocity is of the order c or less, the Newtonian transformation

$$\omega=\omega_1+Uk_1, \qquad k=k_1 \qquad (3)$$

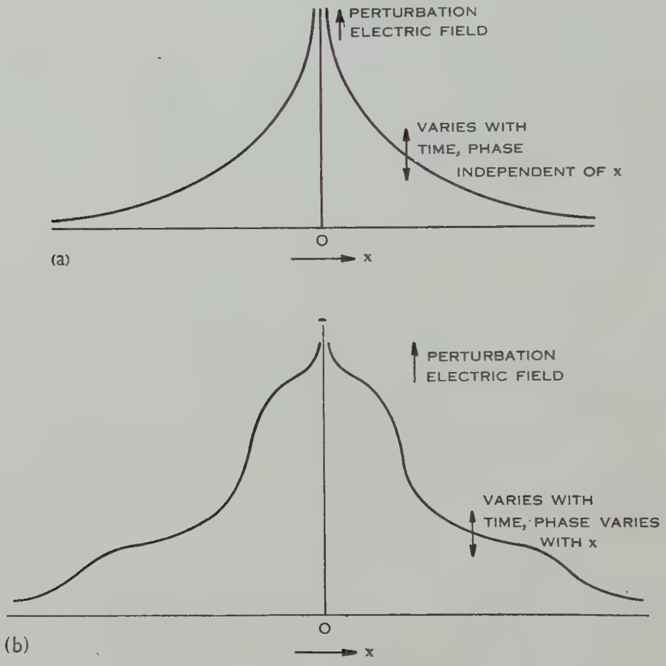


Fig. 1.—Schematic diagrams of pairs of evanescent waves radiated from the emitting surface $x=0$. (*a*) The emitter stationary in the gas : waves have the form $\exp k_i x . \exp i\,\omega_r t$ (k_i positive or negative). (*b*) The emitter moving relative to the gas : waves have the form $\exp(k_i x-\omega_i t).\exp i(\omega_r t-k_r x)$.

may be used. For simplicity this form is adopted in the present discussion, which mainly concerns low gas and wave velocities. This is quite justified, since Newtonian mechanics are assumed by Haeff and by the other authors referred to in Section I in determining electron motions. The Lorentz transformation leads to essentially similar, but more complicated, results. The new dispersion equation, found by combining equations (1) and (3), is

$$k_1=\frac{U\omega_1}{v_t^2-U^2}\left[1\pm\left\{1-\frac{(v_t^2-U^2)(\omega_0^2-\omega_1^2)}{U^2\omega_1^2}\right\}^{\frac{1}{2}}\right]. \qquad (4)$$

When $v_t^2 > U^2$ this equation yields complex values of k_1 for real values of ω_1 in the frequency range

$$\omega_1 < \omega_0\left(1-\frac{U^2}{v_t^2}\right)^{\frac{1}{2}}.$$

Hence, if it were interpreted in the manner of the electron-wave tube theory, it would show travelling waves, one of which grows in space, while remaining steady in time.

This wave growth is spurious ; it does not correspond to an increase of wave energy at the expense of the electron kinetic energy but rather to a process of reflection. As would be expected, the frequency range within which the effect occurs is the same, except for a Doppler shift effect, as that in which a stationary observer sees ordinary evanescent waves.

IV. TRAVELLING EVANESCENT WAVES

It is desirable to consider in more detail the physical nature of the "growing" waves described by equation (4). Later they will be compared with the waves of the electron-wave tube theory. In this theory the emitter moves relative to the gas and it is convenient to make this assumption here. However, contrary to the travelling-wave tube theory, we initially choose an observer stationary in the gas so that equation (1) is applicable.* Since the observer now moves relative to the (steady) emitter he sees a wave whose intensity changes with time so that ω is complex.

Equation (1) may be rewritten

$$v_t^2(k_r+\mathrm{i}k_i)^2=(\omega_r+\mathrm{i}\omega_i)^2-\omega_0^2,$$

and when real and imaginary parts are equated separately :

$$v_t^2(k_r^2-k_i^2)=\omega_r^2-\omega_i^2-\omega_0^2, \qquad (5)$$

$$v_t^2k_rk_i=\omega_r\omega_i. \qquad (6)$$

Solutions are required of the form $\omega_i=Uk_i$ so that equation (6) reduces to

$$v_t^2k_r=U\omega_r, \qquad (7)$$

and equation (5) to

$$k_i^2(v_t^2-U^2)=\omega_0^2-\omega_r^2(1-U^2/v_t^2). \qquad (8)$$

Since all quantities are assumed real, necessary conditions for this form of solution are

$$v_t^2 > U^2 \quad \text{and} \quad \omega_r^2<\omega_0^2(1-U^2/v_t^2).$$

We now have the wave fully specified in terms of the (assumed) quantities ω_r and ω_i.

---

* Clearly the form of the dispersion equation for any given set of circumstances is determined by the movement of the *observer* relative to the gas and is not affected by movement of the *emitter*. Having found the dispersion equation the movement of the emitter is taken account of when more or less arbitrarily choosing two of the four variables ω_r, ω_i, k_r, and k_i. This choice is discussed in Section VI.

The wave is an evanescent wave being reflected back into the emitter. It differs from the stationary evanescent waves illustrated in Figure 1 (*a*) because the emitter is moving. This has two effects: since each successive (in time) intensity maximum occurs in a different place the wave is changed to a travelling wave and also, due to the changing separation of emitter and observer, it grows or decays in time. It may be regarded as a wave packet travelling with *group* velocity $\omega_i/k_i=U$.

The type of wave in which we are particularly interested grows in space while remaining steady in time. Our travelling evanescent wave acquires this property when viewed by an observer moving with velocity U. From equation (3) we determine its parameters* as:

$$\begin{aligned}\omega_1&=\omega_r+\mathrm{i}\omega_i-U(k_r+\mathrm{i}k_i)\\&=\omega_r+Uk_r,\\k_1&=k_r+\mathrm{i}k_i.\end{aligned}$$

Thus the wave has the form:

$$\exp\left(\frac{\omega_0^2}{v_t^2-U^2}-\frac{\omega_r^2}{v_t^2}\right)^{\frac{1}{2}}x\ .\ \exp \mathrm{i}\left\{\left(1+\frac{U^2}{v_t^2}\right)\omega_r t-\frac{U\omega_r}{v_t^2}\ .\ x\right\}.$$

If the rate of growth is not too rapid this wave has a phase velocity

$$\sim\frac{v_t^2+U^2}{U}.$$

A travelling evanescent wave is illustrated schematically in Figure 1 (*b*). For numerical examples we might choose $\omega_r=10^8$ rad sec^{-1}, $v_t=10^7$ cm sec^{-1}, and $U=10^6$ cm sec^{-1}, and two values of ω_0, say 10^8 and 10^9. The two waves are approximately:

$$\exp x\ .\ \exp \mathrm{i}(10^8t-x),$$

and

$$\exp 100x\ .\ \exp \mathrm{i}(10^8t-x).$$

V. The Double Electron Stream

The theory of the electron-wave tube is most fully developed for the case of a double electron stream, the velocities and resonant frequencies of the individual streams being v_a, v_b, ω_a, and ω_b. Haeff (1949*b*) and others have shown that the dispersion equation is

$$\frac{\omega_a^2}{(\omega-v_ak)^2}+\frac{\omega_b^2}{(\omega-v_bk)^2}=1. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

They show that within certain, rather wide limits k has complex values when ω is real and that some of these correspond to waves whose amplitude is increasing in the direction of propagation. Maximum amplification occurs when $\omega_a=\omega_b$ or when $\omega_a/v_a=\omega_b/v_b$.

---

* It will be seen that the transformation of a given wave from one system of axes to another may be effected in either of two ways. The dispersion equation for the new system may be found by a Newtonian transformation and the wave then determined by a suitable substitution of two of the four variables, ω_r, ω_i, k_r, k_i, in this equation. Alternatively the wave parameters, ω, k, may be transformed directly to the new parameters, ω_1, k_1.

As shown above, apparent growth of a wave (complex values of k) may be seen by an observer moving relative to gas in which an evanescent wave is present. The observer for whom the above dispersion equation was derived is moving relative to the gas and so is likely to see such spurious growing waves. To see whether real growing waves are present the dispersion equation for an observer sharing the " mean " velocity of the gas must be derived. It is not, in general, clear what is the " mean " velocity but when the electron streams have equal density ($\omega_a=\omega_b=\omega_0$) then clearly the mean velocity is $\frac{1}{2}(v_a+v_b)$. On transforming to a system with this velocity, equation (**9**) becomes

$$\frac{\omega_0^2}{(\omega_1-vk_1)^2}+\frac{\omega_0^2}{(\omega_1+vk_1)^2}=1, \quad \ldots\ldots\ldots\ldots\ldots \quad \mathbf{(10)}$$

where $v=\frac{1}{2}(v_b-v_a)$. The same result is obtained, of course, by initially assuming two electron streams with velocities $\pm v$. If ω_1 is now assumed real, the values of k_1 are given by

$$v^2k_1^2=\omega_1^2+\omega_0^2\pm(4\omega_1^2\omega_0^2+\omega_0^4)^{\frac{1}{2}}. \quad \ldots\ldots\ldots\ldots \quad \mathbf{(11)}$$

The upper sign corresponds to travelling waves which do not grow or decay and need not be considered further. The lower corresponds, when $\omega_1^2<2\omega_0^2$, to imaginary values of k_1 indicating waves of the form $\exp k_ix \,.\, \exp i\omega_rt$. These are evanescent waves of the type illustrated in Figure 1 (*a*). At no frequency do they grow or decay, so that at no frequency can any electromechanical process occur of the sort envisaged in the electron-wave tube theory.

In the more general case when $\omega_a\neq\omega_b$ it is not clear what is the " mean " velocity of the gas. However, Feinstein and Sen (1951), in their analysis of the two-beam dispersion equation, have shown that amplification (that is, complex values of k) cannot occur when $v_a=-v_b$. This means that if the observer assumes the simple arithmetic mean velocity $\frac{1}{2}(v_a+v_b)$ of the two streams he can never see growing waves. There may be other velocities in which he sees no apparent growth but it is sufficient for our purposes that there should be one. In addition it should be remembered that the case we have analysed ($\omega_a=\omega_b$) is one for which, according to Haeff, growth is strongest.

The real nature of the waves whose growth is predicted by the theory concerned is now fairly clear. They are derived in the same way as the spurious growing waves of Sections III and IV. Starting with a stationary observer, frequency bands are chosen so that evanescent waves are emitted. Observer and emitter are then given a drift velocity and the waves change to a type of travelling wave (since ω_r and k_r are both finite) which is still strongly exponentially attenuated. Of the two modified evanescent waves one then appears to grow and one to decay in space. However, there is no real growth, the observed spatial change in intensity being due to a process of reflection and the wave travel to the fact that the gas carries the wave past the observer.

It is concluded that wave growth in the electron-wave tube by the process hitherto invoked cannot occur. Doubts are raised regarding the somewhat similar theory of the travelling-wave tube and other like electronic devices.

VI. Interpretation of the Dispersion Equation

It would seem that when studying space-charge wave properties that a system of axes should be chosen stationary in the gas. However, even when this is done there may still remain difficulties in the interpretation of the dispersion equation since this does not give either ω or k directly. Two of the four quantities, ω_r, ω_i, k_r, k_i, may be chosen arbitrarily but *the choice must be consistent with physically realizable and relevant conditions.*

The choice really amounts to specifying the method of introduction of the wave into the medium. For example, consider a dispersion equation which is satisfied by complex values of k for real values of ω or alternatively by complex values of ω for real values of k. Both interpretations are reasonable. The first means that the wave is introduced by a steady emitter fixed relative to the observer; the wave then grows (or decays) in space but not in time. The second interpretation is not as obvious but is physically realizable. The wave has the form $\exp -\omega_i t \,.\, \exp i(\omega_r t - k_r x)$ and so grows (if ω_i is negative) in time. Assume that prior to $t=0$ the properties of the medium causing wave growth were non-existent so that the wave was propagated with uniform intensity. When the wave emitted by a distant radiator had permeated the whole gas the growth would then commence. The wave would then appear to grow in time since wave crests passing the observer at later times would have travelled further and so have grown more. An alternative, and perhaps more realistic, way of introducing this wave would be by an emitter whose output increased with time at a rate sufficient to compensate for the growth in space. Again the medium would be permeated by a wave of uniform amplitude which grew in time; again the growth would really be a spatial growth. Thus the alternative interpretations of the dispersion equation are self-consistent, indicating a medium in which the waves grow as they propagate. This would appear to answer Twiss's (1951) main criticism of the electron-wave tube theory.

Not all dispersion equations are as unambiguous as that discussed above. The case of the double electron stream with the observer stationary in the gas raises a difficulty. When ω_1 is assumed real the equation takes the form given in (**11**) and the waves are those discussed above. The alternative assumption that k_1 is real requires consideration; the dispersion equation may be written as

$$\omega_1^2 = \omega_0^2[1 + v^2k_1^2/\omega_0^2 \pm (1 + 4v^2k_1^2/\omega_0^2)^{\frac{1}{2}}]. \qquad \textbf{(12)}$$

When $v^2k_1^2 > 2\omega_0^2$ both waves are travelling waves of constant amplitude. When $v^2k_1^2 < 2\omega_0^2$ the lower sign results in imaginary values of ω_1 so that the waves have the form $\exp -\omega_i t \,.\, \exp -ik_r x$. This is a new type of wave: a non-oscillatory standing wave. It comprises a spatial distribution of electric charge varying sinusoidally throughout the whole of the medium. No oscillations occur, only exponential growth or decay (in time) of the whole pattern.

Bohm and Gross (1949*b*) have used Boltzmann's equation instead of Maxwell's momentum transfer equation to derive the dispersion equation for a pair of equal-density electron beams. One pair of roots of their equation is

$$\omega_1^2 \simeq -Ak_1^2, \qquad \textbf{(13)}$$

where A is a real, positive quantity. This equation is similar to (**12**) (the lower sign chosen) except that ω_1 is imaginary for *all* real values of k_1, that is, for all wavelengths. Bohm and Gross conclude that this dispersion equation indicates an instability of the system and give a " physical reason " for this instability.

The waves might be regarded as limiting cases of travelling growing (or decaying) waves having the form $\exp -\omega_i t \, . \exp i(\omega_r t - k_r x)$. As shown above the dispersion equation for such waves may be interpreted by assuming either ω_1 or k_1 real, corresponding to two different assumed methods of injection each of which is physically realizable. In the limiting case when $\omega_r \to 0$ the wave velocity approaches zero and the wave grows without appreciable movement. There are two objections to the conclusion of Bohm and Gross that the limiting case indicates instability. First, the waves seem physically unrealizable: the limit of zero wave velocity is also a limit of infinite difficulty in injecting the waves. They cannot be propagated into the medium nor instantaneously brought into existence with uniform intensity throughout the medium. In the case of travelling growing waves the alternative assumptions concerning the dispersion equation correspond to alternative (assumed) injection mechanisms but lead to identical wave growths. This raises the second objection to the interpretation of Bohm and Gross: when ω_1 is made real, equations (**12**) and (**13**) provide no evidence of wave growth or medium instability. On the contrary they yield, in the frequency bands concerned, ordinary evanescent waves. Such waves are physically realizable and so provide a satisfactory explanation of the equations. The alternative explanation appears untenable.

The " physical explanation " of the instability is at first sight suggestive but on closer inspection is quite unconvincing. It is not reasonable to assume, as Bohm and Gross have done, that the two beams are velocity modulated (by an initial small disturbance) more or less independently of one another and that the effect is fed back in an amplified form. The electron distribution in each beam contributes equally to a total electric field which, in turn, determines changes in the electron distribution in each beam.

VII. Real Growing Waves

If the mechanism of wave growth of Haeff, Pierce, and other authors is abandoned it is desirable to find an alternative. The proposal depends on electron thermal motions which have been neglected in most of the analyses mentioned but have been included in that of Bohm and Gross (1949*a*, 1949*b*). Before discussing the mechanism of growth it is desirable to mention other effects of thermal motion.

When neither thermal nor drift electron motions are present, then travelling space-charge waves are not possible; the medium may only support stationary oscillations of frequency ω_0. With the introduction of thermal motions a dispersion equation of the form given by equation (**1**) results, and travelling waves are supported at frequencies above ω_0. As seen above, an additional drift motion of the electrons may then cause spurious growing waves to appear at frequencies below ω_0. If differential drift, instead of thermal, motions are introduced then travelling waves are also possible so that the differential drift serves the same purpose as the thermal motions: it diffuses the stationary oscillations and turns them into travelling waves.

The discussion of the previous sections seems to suggest that mass drifts of electrons are never, in themselves, sufficient to cause real wave growth. They may be important, however, in modifying the dispersion equation so as to take better advantage of some primary growth mechanism. There is one such general mechanism which appeals on physically intuitive grounds and may well be the only mechanism capable of causing growth. It depends on the trapping of electrons between potential troughs of the wave and on the extraction of the electron kinetic energy by the wave. The simplest way of extracting the energy is for the wave to slow down, that is, for the phase velocity to decrease. Such an effect requires a medium of variable density and is not likely to be important in the case of the electron-wave tube, although it may play a part in the operation of certain other tubes and is worth considering as a method of amplification.

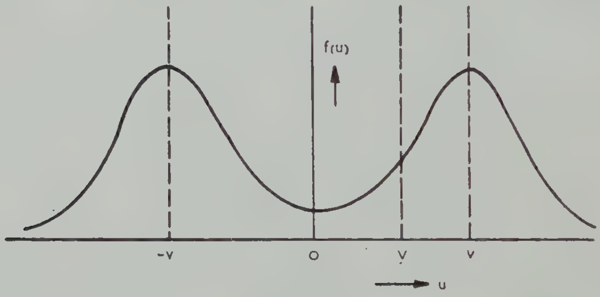


Fig. 2.—A hypothetical velocity distribution in a two-beam electron-wave tube. Velocities are measured relative to the mean electron velocity.

There is another way of extracting the energy, which depends on collisions between electrons, a factor not considered in the simple electron-wave tube theory. It requires that there be electrons close to the phase velocity of the wave; those within a narrow range of velocities will be captured by the wave and oscillate between potential troughs. After a time they will suffer a collision and be scattered back into the statistically steady velocity distribution $f(u)$ of all the electrons. Electrons with initial velocities slightly above the wave velocity will, on an average, lose energy to the wave. Those with velocities slightly below will gain energy from the wave. If there are more of the former than the latter the wave will gain energy and grow, the criterion for growth being $\partial f(u)/\partial u > 0$. The process has been discussed in some detail by Bohm and Gross (1949*b*).*

The possibility of this method of wave growth in the electron-wave tube may be seen by reference to Figure 2. This shows a hypothetical electron velocity distribution relative to the " mean " velocity of all the electrons. The two beams are spread by thermal and other effects but have peaks at velocities $\pm v$. Let the wave velocity be denoted by V; it is given by the possible values of ω_1/k_1 in equation (**10**). Thus we have

---

* They express the growth as an imaginary part of ω. For the present purposes at least, it would be better to express it as an imaginary part of k. This is only a matter of interpreting the dispersion equation appropriately.

$$V = v\{1 + X \pm (4X + X^2)\}^{-\frac{1}{2}}, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{14})$$

where $X = \omega_0^2/\omega_1^2$. Obviously then values of V may occur lying between zero and v and so we have $\partial f(u)/\partial u > 0$ when $u = V$, as shown in the diagram. It will also be clear that the rate of capture of electrons, and hence of wave growth, will be proportional to the intensity of the wave so that the growth is logarithmic as observed.

It is of interest to consider this method of growth for the single-beam electron-wave tube developed by Haeff (1949*b*) and sometimes called the " whistle " tube. The theory of operation of this tube developed by Haeff, along the same lines as for the double-stream tube, has been criticized by Pierce (1950, Ch. 16) who concluded that the fact that the tube worked at all was " something of a mystery ". The tube would work by the new mechanism, provided the electron velocity distribution (about the " mean " velocity) were asymmetrical, as shown in Figure 3. It is then possible for $f(u)$ to have a positive slope in regions corresponding to the velocity of real travelling waves (V).

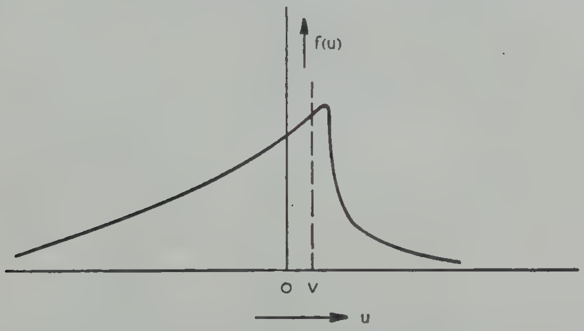


Fig. 3.—A hypothetical velocity distribution (about the mean) in a single-beam electron-wave tube.

Finally, and tentatively, the mechanism is suggested as applying to Pierce's " travelling-wave tube ". The same requirement concerning $f(u)$ is necessary and the function of the helix is to slow down the wave, presumably close to or even below the mean electron velocity, so that the requirement is satisfied.

VIII. DESIGN OF AMPLIFYING TUBES

If the mechanism of space-charge wave growth just described is accepted, there are several rather obvious factors which should be considered when designing actual amplifying devices.

In the first place, the energy gain of the wave increases with the slope of the $f(u)$ curve at the point where $u = V$. The slope should, therefore, be made as steep as possible. The use of two separate beams allows steep slopes to be attained, particularly if the random electron velocities are kept small.

There is a second reason why low electron thermal velocities may be necessary. As Bohm and Gross (1949*a*) have shown, organized oscillations cannot exist at all when the wave phase velocity (ω/k) is below mean thermal velocities. The wave velocity is confined within certain limits by the $f(u)$ distribution so that thermal velocities must be kept within closer limits.*

A second factor promoting growth is the frequency of collisions of electrons with electrons. This should be as high as possible so that the rate of circulation of electrons between the trapped and untrapped conditions is a maximum. Once again a reduction in thermal motions would probably help since the effect of higher electron speed on collision frequency is more than offset by the change in collision cross section. The effect of electron collisions with heavy ions or atoms is to damp the wave by a scattering process.

A desirable factor in addition to gain is bandwidth and these requirements will inevitably clash with one another. Thus change of frequency will change X in equation (**14**) and hence V. A large bandwidth requires a large velocity range over which $f(u)$ is large and positive. The combination of a steep and long slope of the $f(u)$ curve implies a large value of total electron density and this in turn is limited by the minimum frequency to be used.

It seems likely that the general form of the dispersion equation for several beams of different densities will be

$$\omega/k=F(v_1, v_2, \ldots, X_1, X_2, \ldots). \quad \ldots\ldots\ldots\ldots \quad (\mathbf{15})$$

By similar choice of the various parameters an optimum balance between the different requirements may be effected. This illustrates the importance of introducing a second electron beam (or alternatively, perhaps, a helix) ; in effect it provides another degree of freedom which can be used to obtain optimum results.

Finally, the possibility suggests itself of using a wave whose velocity (relative to the mean electron velocity) decreases as it propagates. This would allow a large proportion of the kinetic energy of the electrons (within the range of variation of V) to be abstracted. A method of achieving this effect might be by using a magnetic field which is not constant in direction. This would not only change the X terms in equation (**15**) but introduce fresh terms depending on the strength and direction of the magnetic field.

IX. CONCLUSIONS

(1) The current theory of the electron-wave tube is untenable ; the wave growth predicted is spurious, owing to the movement of the observer relative to the gas which carries the wave. The criticism applies to some other theories of growing waves including, perhaps, that of the travelling-wave tube.

---

* The converse argument is that, having a certain, inevitable spread of thermal velocities the spread of *ordered* velocities must be greater than a certain minimum value if coherent amplified waves are to occur. This may explain the observed wide spread of ordered velocities in practical amplifying devices.

(2) Theories of the type concerned may be correctly developed without reference to boundaries, although if these are present their effects must be included as additional factors.

(3) Two rules must be observed in using the method of substitution analysis to study waves in gases. First, the frame of reference in which the equations are developed should be stationary relative to the gas. Second, when (more or less arbitrarily) choosing real or imaginary parts of frequency or propagation constant in the dispersion equation, the choice must be consistent with physically realizable and relevant conditions.

(4) The only known method of growth of space-charge waves is by the trapped electron mechanism, by which electrons are trapped between potential troughs and their kinetic energy abstracted. This seems the most likely explanation of the amplification which undoubtedly does occur in the electron-wave and other growing-wave tubes.

X. Acknowledgment

The author wishes to thank Mr. R. F. Mullaly for several most helpful discussions.

XI. References

Bailey, V. A. (1948).—*Aust. J. Sci. Res.* A **1** : 351.
Bailey, V. A. (1950).—*Phys. Rev.* **78** : 428.
Bailey, V. A. (1951).—*Phys. Rev.* **83** : 439.
Bohm, D., and Gross, E. P. (1949*a*).—*Phys. Rev.* **75** : 1851.
Bohm, D., and Gross, E. P. (1949*b*).—*Phys. Rev.* **75** : 1864.
Feinstein, J., and Sen, H. K. (1951).—*Phys. Rev.* **83** : 405.
Haeff, A. V. (1948).—*Phys. Rev.* **74** : 1532.
Haeff, A. V. (1949*a*).—*Phys. Rev.* **75** : 1546.
Haeff, A. V. (1949*b*).—*Proc. Inst. Radio Engrs., N.Y.* **37** : 4.
Nergaard, L. S. (1948).—*R.C.A. Rev.* **9** : 585.
Piddington, J. H. (1955).—Growing electromagnetic waves. *Phys. Rev.* (in press).
Pierce, J. R. (1948).—*J. Appl. Phys.* **19** : 231.
Pierce, J. R. (1949).—*Proc. Inst. Radio Engrs., N.Y.* **37** : 980.
Pierce, J. R. (1950).—" Travelling Wave Tubes." (Van Nostrand : New York.)
Pierce, J. R., and Herbenstreit, W. B. (1948).—*Bell Syst. Tech. J.* **28** : 33.
Rydbeck, O. E. H., and Forsgren, S. K. H. (1951).—Chalmers Tek. Högsk. Handl. No. 102.
Thomson, J. J., and Thomson, G. P. (1933).—" Conduction of Electricity in Gases." Vol. 2, p. 353. (Cambridge Univ. Press.)
Tonks, L., and Langmuir, I. (1929).—*Phys. Rev.* **33** : 195.
Twiss, R. Q. (1951).—*Proc. Phys. Soc. Lond.* B **64** : 654.

FORMULAE FOR THE MEAN LOSSES OF ENERGY IN COLLISIONS OF SLOW ELECTRONS MOVING IN DIATOMIC GASES

By L. G. H. Huxley*

[*Manuscript received September* 15, 1955]

Summary

Formulae are derived for the mean losses of energy in collisions with molecules by electrons moving in a steady state of motion in a diatomic gas in an electric field. Although the theory is incomplete and crude, yet the formulae describe the experimental measurements closely in the smaller ranges of electronic energy.

Such formulae find application in the theory of radio wave interaction.

I. Introduction

Although the results of the measurements of the mean losses of energy experienced by slow electrons in a steady state of motion among the molecules of diatomic gases have been available for some time, there do not appear to exist any formulae, with a reasonable physical basis, that represent the experimental results over any significant range of the mean energies of the electrons. Recently, the problem has been treated by the methods of quantum mechanics by Gerjuoy and Stein (1954, 1955) and it has also been discussed by Massey and Burhop (1952).

Because such formulae would be of practical value, especially in situations inaccessible to direct laboratory measurement (as, for instance, when the mean energies of agitation of the electrons do not differ greatly from that of the gas molecules), the aim of the investigation that follows is to seek satisfactory formulae by means of crude theory, rather than to present a rigorous theory of the mechanism of the losses of energy.

Recent measurements of the diffusion of slow electrons in deuterium made in this department by Miss B. I. H. Hall (1955), are significant in this context. It was found that the collisional cross sections of the molecules in collisions with electrons moving with any given speed U were the same in deuterium and in hydrogen, as would be expected from the identity of the electronic shells of their molecules. However, the mean energy ΔQ_H lost by an electron with energy Q in collision with a molecule of hydrogen is twice the mean energy ΔQ_D lost by an electron with the same energy Q in collisions with molecules of deuterium, for all values of the energy Q. Since the ratio of the mean losses of energy $\Delta Q_H/\Delta Q_D$ is the same as the inverse ratio M_D/M_H of the masses of molecules, it is suggested that the only kinds of exchanges of energy in collisions concerned are those that are inversely proportional to the masses of the molecules. Two types of collision

---

* Department of Physics, The University of Adelaide.

that exhibit this behaviour are perfectly elastic collisions between electrons and molecules and collisions that produce changes in the rotational states of the molecules.

II. Perfectly Elastic Collisions

Were electrons with energy Q to collide with molecules with energy Q_0 as if each were a perfectly elastic sphere then the mean energy lost by an electron in a collision, when $Q>Q_0$ and ratio of their masses $m/M \ll 1$, is

$$\Delta Q_e=(2m/M)(Q-Q_0), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

and the proportion of its energy lost in a collision is, on the average,

$$\eta=\Delta Q_e/Q=2m/M(1-1/k_T), \quad \ldots\ldots\ldots\ldots \quad (2)$$

where $k_T=Q/Q_0$ (Townsend's energy factor).
Also

$$\eta k_T=Q_e/Q_0=2m/M(k_T-1). \quad \ldots\ldots\ldots\ldots\ldots \quad (3)$$

If M is the mass of a molecule of hydrogen then $2m/M=5\cdot 44\times 10^{-4}$.

III. Rotational States in Hydrogen

The energies ε_J associated with the rotational states of a diatomic molecule are

$$\varepsilon_J=(h^2/8\pi^2 I)J(J+1), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

where h is Planck's constant, I the moment of inertia of the molecule, and J an integer.

The rotational angular momentum of the molecule in state J is

$$P_J=\{J(J+1)\}^{\frac{1}{2}}h/2\pi. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

The relative abundances p_J of molecules of hydrogen in the various states J are:

parahydrogen ($J=0, 2, 4, \ldots$)

$$p_J=(2J+1)\exp(-\varepsilon_J/kT),$$

orthohydrogen ($J=1, 3, 5, \ldots$)

$$p_J=3(2J+1)\exp(-\varepsilon_J/kT),$$

where k is Boltzmann's constant.

Table 1

RELATIVE POPULATIONS OF STATES IN HYDROGEN

Molecule	Parahydrogen			Orthohydrogen		
J	0	2	4	1	3	5
Relative population ..	1	0·825	$2\cdot 4\times 10^{-2}$	4·95	0·57	$5\cdot 4\times 10^{-3}$

When $T=288$ °K; $h^2/(8\pi^2 I)\simeq Q_0/5$; $Q_0=3/2$. $kT=5\cdot 96\times 10^{-14}$ ergs.
The relative populations of the states with J small at $T=288$ °K are therefore the values given in Table 1.

Energy steps

$$\Delta\varepsilon_{J'J}=\varepsilon_{J'}-\varepsilon_{J}=(J'-J)(J'+J+1)\times 0\cdot 198Q_0.$$

Whence

$$\varepsilon_{10}=0\cdot 39Q_0\,;\quad \varepsilon_{21}=0\cdot 79Q_0\,;\quad \varepsilon_{32}=1\cdot 19Q_0\,;\quad \varepsilon_{43}=1\cdot 58Q_0.$$

Steps of angular momentum

$$\Delta P_{J'J}=P_{J'}-P_{J}=[\{J'(J'+1)\}^{\frac{1}{2}}-\{J(J+1)\}^{\frac{1}{2}}]h/2\pi$$

$$P_{10}=1\cdot 49\times 10^{-27}\,;\quad P_{21}=1\cdot 08\times 10^{-27}\,;\quad P_{32}=1\cdot 06\times 10^{-27}\,;\quad P_{43}=1\cdot 06\times 10^{-27}$$

$$P_{20}=2\cdot 56\times 10^{-27}\,;\quad P_{31}=2\cdot 15\times 10^{-27}\,;\quad P_{42}=2\cdot 12\times 10^{-27}\,;\quad P_{53}=2\cdot 11\times 10^{-27}$$

erg sec.

IV. Necessary Conditions for Production of a Transition $\Delta P_{J'J}$ $(J'>J)$

These are:

(*a*) $Q\geqslant\Delta\varepsilon_{J'J}$;

(*b*) the angular momentum of the colliding electron about the centre of the molecule when approaching it must equal or exceed $\Delta P_{J'J}$.

In practice, if condition (*b*) is fulfilled then so also is condition (*a*).

Suppose that when the impact parameter r_i of the electron's orbit about the molecular centre exceeds a critical distance r_c the mutual interaction of the molecule and the electron has vanished so that the change $\Delta P_{J'J}$ cannot take place. It is therefore necessary that the speed U of an electron should exceed U_c if the transference of momentum $\Delta P_{J'J}$ is to be possible, where $U_c=\Delta P_{J'J}/mr_c$.

If the speed U of the electron exceeds U_c then the increment $\Delta P_{J'J}$ is possible provided $r\leqslant r_i\leqslant r_c$ where $r=\Delta P_{J'J}/mU$. Let the probability that an electron of speed U should produce the increment $\Delta P_{J'J}$ in the angular momentum when these conditions are fulfilled be $p(U)$.

V. The Mean Energy Lost by an Electron in a Collision

Let the speeds U of the electrons in collisions with molecules be distributed according to some law that the proportion of speeds that exceed U but do not exceed $(U+\mathrm{d}U)$ is $f(U)\mathrm{d}U$. The total number of collisions per second per electron in which the conditions for the possibility of a transference of angular momentum $\Delta P_{J'J}$ are satisfied is

$$N=n\pi r_c^2\int_{U_c}^{\infty}(1-r^2/r_c^2)Uf(U)\,\mathrm{d}U \quad \ldots\ldots\ldots\ldots \quad (6)$$

$$=n\pi r_c^2 I,$$

where n is the number of molecules in unit volume, $U_c=\Delta P_{J'J}/mr_c$, and $r=\Delta P_{J'J}/mU$. Also $r/r_c=U_c/U$ and I is the integral.

The total loss of energy in these collisions is

$$n\pi r_c^2\Delta\varepsilon_{J'J}\int_{U_c}^{\infty}p(U)(1-r^2/r_c^2)Uf(U)\,\mathrm{d}U=n\bar{p}A_c\Delta\varepsilon_{J'J}I,$$

where $\bar{p}$ is the mean value of the probability $p(U)$ averaged over all speeds $U > U_c$; and $A_c = \pi r_c^2$.

This total loss of energy averaged over all collisions of this type contributes to the mean energy lost by an electron in any collision the amount

$$\Delta Q_{J'J} = (\bar{p}A_c/A)(\Delta\varepsilon_{J'J}/\bar{U})I, \quad \ldots\ldots\ldots\ldots\ldots \quad (7)$$

where A is the mean collisional cross section of the molecules when the mean speed of the electrons is $\bar{U} = \int_0^\infty Uf(U)\mathrm{d}U$.

The total mean energy lost by an electron in a collision taking into account elastic and rotational losses is

$$\Delta Q = \Delta Q_e + \Sigma_{J'J}\Delta Q_{J'J}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

VI. Special Distributions $f(U)$

It is convenient to obtain the formula for $\Delta Q_{J'J}$ in the special cases in which $f(U)$ corresponds respectively to laws of distribution of Maxwell and Druyvesteyn.

(a) *Maxwell's Distribution*

$$f(U) = (4/\alpha^3\sqrt{\pi})U^2 \exp(-U^2/\alpha^2), \quad \bar{U} = 2\alpha/\sqrt{\pi},$$

where α is the most probable speed.

The integral in equation (6) becomes

$$I = I_1 - I_2,$$

where

$$I_1 = 2\alpha/\sqrt{\pi}\int_{(U_c/\alpha)^2}^{\infty} \mathrm{e}^{-y}y\mathrm{d}y = (2\alpha/\sqrt{\pi})(1 + U_c^2/\alpha^2)\exp(-U_c^2/\alpha^2),$$

$$I_2 = \frac{2U_c^2}{\alpha\sqrt{\pi}}\int_{(U_c^2/\alpha^2)}^{\infty} \mathrm{e}^{-y}\mathrm{d}y = (2U_c^2/\alpha\sqrt{\pi})\exp(-U_c^2/\alpha^2).$$

Whence

$$I = I_1 - I_2 = (2\alpha/\sqrt{\pi})\exp(-U_c^2/\alpha^2) = \bar{U}\exp(-U_c^2/\alpha^2).$$

It follows from equation (7) that at $T = 288$ °K,

$$\begin{aligned}\Delta Q_{J'J} &= (\bar{p}A_c/A)\Delta\varepsilon_{J'J}\exp(-U_c^2/\alpha^2)\\ &= (\bar{p}A_c/A)(J'-J)(J'+J+1)(0{\cdot}198Q_0)\exp(-U_c^2/\alpha^2). \quad \ldots\ldots \quad (9)\end{aligned}$$

The contribution of this source of loss to the total mean proportional loss of energy in a collision is

$$\left.\begin{aligned}\eta_{J'J} &= \Delta Q_{J'J}/Q = \Delta Q_{J'J}/(Q_0k_T)\\ &= (a/k_T)\exp(-U_c^2/\alpha^2),\\ \text{where} \quad a &= 0{\cdot}198\bar{p}A_c(J'-J)(J'+J+1)/A.\end{aligned}\right\} \quad \ldots\ldots \quad (10)$$

Since $U_c = \Delta P_{J'J}/mr_c = Ch/2\pi mr_c = 1\cdot 05 \times 10^{-27} C/mr_c$ and $Q = Q_0 k_T = 3m\alpha^2/4$, where $C = \{J'(J'+1)\}^{\frac{1}{2}} - \{J(J+1)\}^{\frac{1}{2}}$ and $m = 0\cdot 912 \times 10^{-27}$ g, it follows that, when the temperature $T = 288$ °K,

$$\left.\begin{aligned} U_c^2/\alpha^2 &= \frac{1\cdot 50 \times 10^{-14} C^2}{r_c^2 k_T} = b/k_T, \\ \text{where} \quad b &= 1\cdot 50 \times 10^{-14} C^2/r_c^2. \end{aligned}\right\} \quad \dots\dots\dots \quad (\mathbf{11})$$

Consequently, equation (**10**) may be written

$$\eta_{J'J} = (a/k_T) \exp(-b/k_T). \quad \dots\dots\dots\dots\dots\dots \quad (\mathbf{12})$$

Also

$$\eta_{J'J} k_T = \Delta Q_{J'J}/Q_0 = a \exp(-b/k_T). \quad \dots\dots\dots\dots \quad (\mathbf{13})$$

The measured value of $\eta_{J'J} k_T$ is not represented by this formula for there exist inverse transitions in which electrons receive energy from the molecules. These are such that in thermal equilibrium with $k_T = 1$, the effective loss of energy in collisions is, on the average, zero. The effective values of $\eta_{J'J} k_T$ are therefore assumed to be

$$\left.\begin{aligned} \eta_{J'J} k_T &= a[\exp(-b/k_T) - \exp(-b)], \\ \text{and} \quad \eta_{J'J} &= (a/k_T)[\exp(-b/k_T) - \exp(-b)]. \end{aligned}\right\} \quad \dots\dots \quad (\mathbf{14})$$

These are then summed over all transitions $J \rightarrow J'$ to give the total effective losses $\eta_r k_T$ in collisions that arise from rotational transitions, thus,

$$\eta_r k_T = k_T \Sigma \eta_{J'J} = \Sigma a_{J'J} \exp[(-b_{J'J}/k_T) - \exp(-b_{J'J})]. \quad \dots\dots \quad (\mathbf{15})$$

(*b*) *Druyvesteyn's Distribution*

$$f(U) = \frac{4U^2}{\alpha^3 \Gamma(3/4)} \exp(-U^4/\alpha^4).$$

Then

$$\overline{U} = \alpha/\Gamma(3/4) = 0\cdot 816\alpha, \quad \overline{U^2} = \alpha^2 \Gamma(5/4)/\Gamma(3/4) = 0\cdot 75\alpha^2,$$

whence $\alpha^2 = 2\cdot 7Q/m$. The most probable speed is $\alpha/\sqrt{2}$. It follows that

$$U_c^2/\alpha^2 = 75 \times 10^{-16} (C^2/r_c^2 k_T), \quad \dots\dots\dots\dots\dots \quad (\mathbf{16})$$

where C is defined as in equation (**11**).

It may be shown that the formulae that here correspond to equation (**14**) are

$$\left.\begin{aligned} \eta_{J'J} k_T &= c[F(d/k_T) - F(d)], \\ \text{where} \quad F(d/k_T) &= 2/\sqrt{\pi} \exp(-d^2/k_T{}^2) - 2(d/k_T) \operatorname{erfc}(d^2/k_T{}^2), \end{aligned}\right\} \quad \dots \quad (\mathbf{17})$$

$$d = 75 \times 10^{-16} (C/r_c)^2, \quad c = 0\cdot 198 \sqrt{(\tfrac{1}{2}\pi)} \bar{p} A_c (J' - J)(J' + J + 1)/A.$$

The coefficients a and c that appear in formulae (**14**) and (**17**) are, in general, not constants, but are functions of k_T through the term $\bar{p}/A$ that appears in equation (**7**) so that a and c would be more correctly given in the form, constant $\times\varphi(k_T)/\psi(k_T)$. Although the experimental dependence of A upon k_T is known for some gases there is no information about $\bar{p}(k_T)$. In what follows it is supposed that $\bar{p}(k_T)$ is in effect a constant.

VII. ENERGY LOSSES OF ELECTRONS IN HYDROGEN WHEN $k_T<5$

According to Section II, the sum of the relative populations of the states $J=1$, 2, and 3 is $6\cdot35$, whereas that of the state $J=0$ is unity. Moreover, the increments of angular momentum ΔP_{12}, ΔP_{23}, and ΔP_{34} are about $1\cdot07\times10^{-27}$ erg sec, whereas ΔP_{01} is appreciably greater with a value of $1\cdot49\times10^{-27}$ erg sec.

It is to be expected therefore that, as k_T is increased from $k_T=1$ (thermal equilibrium) to, say, $k_T=5$, the electrons lose energy chiefly in exciting the transitions $P_1\rightarrow P_2$, $P_2\rightarrow P_3$, $P_3\rightarrow P_4$. It is reasonable to group these transitions and to assign to them a common value of $r_cU_c=1\cdot07\times10^{-27}/m$ erg sec.

In hydrogen the collisional cross section A is closely represented by $A=0\cdot816\,(10\cdot7+k_T)\times10^{-16}$ cm², when $1\cdot73<k_T<5\cdot1$. Consequently the following formula is suggested for η_r:

$$\eta_r=\frac{\alpha}{k_T(10\cdot7+k_T)}\,[\exp\,(-b/k_T)-\exp\,(-b)]. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{18})$$

η_r is a maximum where

$$b/k_T=[1+k_T(10\cdot7+k_T)][1-\exp\,\{-b(1-1/k_T)\}]. \quad \ldots\ldots\ldots \quad (\mathbf{19})$$

A first approximation to b is $[1+k_T/(10\cdot7+k_T)]k_T$. A second and adequate approximation is

$$b=k_T[1+k_T/(10\cdot7+k_T)][1-\exp\,\{-(1+k_T/(10\cdot7+k_T))(k_T-1)\}]. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{20})$$

When b is found from equation (**20**), α may be found from the experimental value of $\eta_{\text{max.}}$. In hydrogen when $1<k_T<5$ it was found that $\alpha=0\cdot271$; $b=4\cdot49$. A comparison is made in Table 2 between η_r calculated from equation (**18**) with these values of α and b and the experimental results of Crompton and Sutton (1952). The figures in the second row are the measured values of ηk_T and the third row gives the corresponding values of $(\eta k_T)_e$, the contribution to ηk_T of the losses in elastic

TABLE 2

COMPARISON OF EXPERIMENTAL AND CALCULATED VALUES OF η_r

k_T	1·34	1·73	2·5	3·4	4·2	5·1	5·9	6·8
$\eta k_T\times10^4$	7·16	18·3	41·2	62·2	78·0	95·3	114	129
$(\eta k_T)_e\times10^4$	1·85	3·96	8·16	13·1	17·4	22·3	26·8	35·8
$(\eta k_T)_r\times10^4$	5·31	14·4	33·0	49·1	60·6	73·0	84·2	84·2
$\eta_r\times10^4$	3·98	8·25	13·2	14·4	14·4	14·3	14·3	14·3
$\eta_r\times10^4$ (calc.) ..	4·05	8·30	12·7	14·4	14·4	14·3	12·7	11·5

collisions (Section II). The figures in the fourth row are the result of subtracting the third row from the second and represent $(\eta k_T)_r$, the contribution of the rotational losses. The fifth row is η_r. The last row contains the values of η_r calculated from equation (**18**). Agreement is good within the range $1<k_T<5$. When k_T exceeds 5 the formula fails because additional sources of loss of energy become important and because A increases more slowly with k_T than is suggested by the formula $A=0\cdot816\,(10\cdot7+k_T)\times10^{-16}$ cm³. The appropriate adaptations of formulae (**21**) based on Druyvesteyn's distribution do not describe the observations as accurately as does formula (**18**). The critical radius r_c is, according to equation (**11**), given by $r_c=(1\cdot52\times10^{-14}/b)^{\frac{1}{2}}C$. In transitions $J'\rightarrow J=1$; $J=1, 2, 3$, the value of C is about $1\cdot02$. With $b=4\cdot49$ the value of r_c is $5\cdot6\times10^{-8}$ cm. If it is supposed that the appropriate transitions are $J\rightarrow J'=J+2$, then the value of r_c is approximately twice as great.

VIII. ENERGY LOSSES IN OXYGEN

Although the assumption that the distribution function $f(U)$ is that of Maxwell appears to be satisfactory for the motion of electrons in hydrogen when $k_T<5$, yet it is to be anticipated that Druyvesteyn's distribution function is more nearly correct at larger values of k_T. This expectation is confirmed by an analysis of some unpublished measurements by R. W. Crompton and D. J. Sutton in this department of the motions of electrons in oxygen in which the variation of A with k_T is small. It is found that the dependence of η upon k_T is very closely described by formula (**17**) based on Druyvesteyn's function and not at all well by formula (**14**) based on Maxwell's, over the range of values of k_T from 8 to 30, the curve of η *v.* k_T in oxygen differs markedly from those of hydrogen and nitrogen in that the first maximum is clearly marked and lies near $k_T=14\cdot4$ instead of $k_T=4$. A consequence of this displacement of the maximum to a relatively large value of k_T is that the constant term $F(d)$ in equation (**17**) is negligibly small. Thus in oxygen

$$\left.\begin{aligned} &\eta_r=(c/k_T)F(d/k_T)=(c/d)xF(x),\\ \text{where}\qquad &x=d/k_T,\\ \text{and}\qquad &F(x)=(2/\sqrt{\pi})\exp(-x^2)-2x\operatorname{erfc}(x^2). \end{aligned}\right\}\quad\ldots\ldots\ldots\ (\mathbf{21})$$

The function $xF(x)$ possesses a maximum value $2\cdot03$ at $x=0\cdot435$, consequently, if the experimental maximum is $(\eta_r)_{\text{max.}}$ at $(k_T)_{\text{max.}}$, then, $d=0\cdot435(k_T)_{\text{max.}}$ and $c/d=(\eta_r)_{\text{max.}}/2\cdot03$.

The large value of $(k_T)_{\text{max.}}$ in oxygen is to be attributed to the fact that, because of the absence of nuclear spin, only the rotational states corresponding to $J=1, 3, 5, \ldots$ exist, the most populous states at $T=288$ °K being those for which J is of the order of 8. The smallest of the increments of angular momentum require transitions $J\rightarrow J+2$, and when J is as large as 8 all $\Delta P_{J,(J+2)}\rightarrow 2h/2\pi=2\cdot102\times10^{-27}$ erg sec. Thus, in formula (**17**), $C=2$ and

$$r_c^2=75\times10^{-16}C^2/d=3\times10^{-14}/0\cdot435(k_T)_{\text{max.}}\simeq4\cdot8\times10^{-15},$$
$$r_c\simeq7\times10^{-8}\text{ cm.}$$

The particular form of equation (**21**) that describes the proportional energy loss in oxygen is :

$$\eta = 4\cdot 17 \times 10^{-2}\,(6\cdot 25/k_T)F(6\cdot 25/k_T). \quad \ldots\ldots\ldots\ (\mathbf{22})$$

Table 3 gives a comparison of the predictions of this formula and the measured values of η of Crompton and Sutton, on the assumption that the distribution function $f(U)$ is that of Druyvesteyn.

TABLE 3

COMPARISON OF PREDICTED AND MEASURED VALUES OF η

Small Values of k_T

k_{TD}	4·1	5·3	6·2	7·2
$10^3 \times \eta_D$	3·09	4·14	5·10	5·75
$10^3 \times \eta$ (calc.) . .	1·04	2·75	4·17	5·29

Large Values of k_T

k_T	7·9	8·8	9·5	10·1	11·0	12·7	15·8	19·4	22·3	36·0
$10^3 \times \eta$ . .	6·45	6·92	7·30	7·7	8·02	8·35	8·35	8·03	7·70	6·40
$10^3 \times \eta$ (calc.)	6·2	6·90	7·35	7·7	8·0	8·34	8·36	8·08	7·66	5·0

It can be seen that the agreement between the calculated and measured values is close when $8 < k_T < 30$ but that divergences occur outside these limits. At small values of k_T the distribution function $f(U)$ is presumably tending towards that of Maxwell so that formula (**22**) is not applicable. The experimental curve for η *v.* k_T exhibits a minimum at $k_T = 36$ with a progressive rise as k_T increases. There is therefore an additional process producing loss of energy whose effects are not taken account of by formula (**22**).

IX. ENERGY LOSSES IN NITROGEN

It was found (Crompton, Huxley, and Sutton 1953) that in nitrogen and air the collisional cross section A is proportional to the mean speed $\overline{U}$ of the electrons and therefore to k_T. In nitrogen, when $k_T < 9$, $A = 2\cdot 86 k_T^{\frac{1}{2}} \times 10^{-16}$. In this gas, therefore, it is to be expected that η would be represented by a formula of the form

$$\eta = \frac{\alpha}{k_T{}^{3/2}}\,[\exp(-b/k_T) - \exp(-b)]. \quad \ldots\ldots\ldots\ (\mathbf{23})$$

Because measurements of η are lacking for values of k_T less than that at which η attains its maximum value in the range of small values of k_T it is not possible to ascribe values to α and b with any confidence.

X. THE MEAN RATE R AT WHICH AN ELECTRON LOSES ENERGY

The mean energy effectively lost by an electron in unit time in producing rotational changes follows from equation (**6**) and the following equation.

Since in both the Maxwell and Druyvesteyn forms of the distribution function the mean speed $\overline{U}$ appears as a factor in the expression for the total energy lost per second, this rate of loss may be written in the form

$$\left.\begin{aligned} &R/n=\alpha Q^{\frac{1}{2}}\,[\exp(-\beta/Q)-\exp(-\beta/Q_0)], && \text{Maxwell's distribution} \\ \text{or}\quad &R/n=\gamma Q^{\frac{1}{2}}[F(\delta/Q)-F(\delta/Q_0)], && \text{Druyvesteyn's distribution} \end{aligned}\right\} \quad (\mathbf{24})$$

where α, β, γ, and δ are constants independent of Q and Q_0 and the function F is defined in equation (**17**).

The forms of these expressions that are adapted to the laboratory measurements at $T=288$ °K are

$$\left.\begin{aligned} &R/n=ak_T^{\frac{1}{2}}\,[\exp(-b/k_T)-\exp(-b)], \\ \text{and}\quad &R/n=ck_T^{\frac{1}{2}}\,[F(d/k_T)-F(d)], \end{aligned}\right\} \quad \ldots\ldots\ldots (\mathbf{25})$$

where a and c are constants, and $b=\beta/Q_{288}$, $d=\delta/Q_{288}$, and $\delta=0{\cdot}435\beta$. Q_{288} is Q_0 at $T=288$.

The quantity R/n is doubly convenient. First, it is independent of the mean collisional cross section A; secondly, it is immediately derivable from the laboratory measurements of the drift speeds W of electrons in a steady state of motion through a gas in a steady electric field Z.

In terms of the drift speed W and electronic charge e, R is given by $R=ZeW$. Let p be the pressure of the gas in millimetres of mercury at $T=288$ °K, and let n be the number of molecules of the gas in a cubic centimetre.

Then $n=3{\cdot}35\times10^{16}p$ and it follows that

$$(3{\cdot}35\times10^{16}\times300/4{\cdot}8\times10^{-10})R/n=2{\cdot}09\times10^{28}R/n=(Z/p)W, \quad \ldots\ldots (\mathbf{26})$$

where Z is expressed in volts per centimetre.

Since k_T is a function of Z/p it follows that both R/n and $(Z/p)W$ may be found as experimental functions of k_T and the result may be compared with the theoretical formulae (**25**). If equations (**26**) correspond to reality, then the experimental quantity $W(Z/p)k_T^{-\frac{1}{2}}$ should be related to k_T, when $T=288°$ (the laboratory temperature) as follows:

$$\left.\begin{aligned} &W(Z/p)k_T^{-\frac{1}{2}}=\text{constant }[\exp(-b/k_T)-\exp(-b)], && \text{Maxwell's distribution} \\ &W(Z/p)k_T^{-\frac{1}{2}}=\text{constant }[F(d/k_T)-F(d)]. && \text{Druyvesteyn's distribution} \end{aligned}\right\} \ldots (\mathbf{27})$$

It is convenient to test these expressions against the measurements of electronic motions in oxygen because in this gas, not only are the terms $\exp(-b)$ and $F(d)$ negligible but a single type of rotational loss is present over a wide range of the parameter k_T (Section VIII).

Table 4 shows that good agreement is obtained with the first of formulae (**27**) when $k_T<9$, and with the second when $10<k_T<30$.

It would appear therefore, that the transition from Maxwell's law of distribution to that of Druyvesteyn takes place in the range of values of k_T $8<k_T<10$.

Table 4

COMPARISON OF CALCULATED AND OBSERVED VALUES OF y

Smaller values of k_T; $y=(Z/p)Wk_T^{-\frac{1}{2}}=1\cdot55\times10^6\exp(-14\cdot4/k_T)$

k_T (Maxwell)	4·7	6·0	7·1	8·2	9	10
(y observed) $\times10^{-4}$ ..	7·4	13·0	19·9	26·1	32·3	39·1
(y calculated) $\times10^{-4}$	7·1	14·0	20·3	26·6	31·2	36·6

Larger values of k_T; $y=(Z/p)Wk_T^{-\frac{1}{2}}=1\cdot94\times10^6F(6\cdot25/k_T)$

k_T (Druyvesteyn) ..	8·8	9·5	10·1	11·0	12·7	15·8	19·4	22·3	36·0
(y observed) $\times10^{-5}$ ..	4·2	4·91	5·72	6·42	7·86	9·89	11·6	12·6	17·5
(y calculated) $\times10^{-5}$	4·57	5·20	5·80	6·48	7·86	9·89	11·6	12·4	15·9

The particular forms assumed by the expressions (**24**), as derived from the experimental measurements at $T=288$ are:

$$\left.\begin{aligned} R/n&=3\cdot04\times10^{-16}Q^{\frac{1}{2}}\exp(-8\cdot58\times10^{-13}/Q), \quad k_T<8,\\ R/n&=3\cdot80\times10^{-16}Q^{\frac{1}{2}}F(-3\cdot72\times10^{-13}/Q), \quad 10<k_T<30. \end{aligned}\right\} \quad \ldots\ldots (\mathbf{28})$$

In equation (**9**) write $R=\nu\Delta Q_{J'J}$ and put $\nu=nA\bar{U}$, $\bar{U}=4\cdot33\times10^{13}Q^{\frac{1}{2}}$, $J'-J=2$, $J=8$, $(\Delta\varepsilon_{J'J})_{\text{oxygen}}=0\cdot198Q_0/16$. Comparison with equation (**28**) suggests that $\bar{p}=3\times10^{-2}$.

The quantity R for air is of fundamental importance in the theory of radio wave interaction, but this aspect of the study is discussed in another paper.

XI. Acknowledgments

The author is indebted to Dr. R. W. Crompton and Miss B. I. H. Hall for criticisms and discussions of this paper; and to Drs. R. W. Crompton and D. J. Sutton for making available their unpublished measurements of the losses of energy of electrons in oxygen. He is also indebted to Professors H. S. W. Massey and H. S. Green and to Dr. O. Bergmann for discussions on the subject-matter of the paper.

XII. References

Crompton, R. W., and Sutton, D. J. (1952).—*Proc. Roy. Soc.* A **215**: 467.

Crompton, R. W., Huxley, L. G. H., and Sutton, D. J. (1953).—*Proc. Roy. Soc.* A **218**: 507.

Gerjuoy, E., and Stein, S. (1954).—Westinghouse Research Laboratories Sci. Res. Pap. 60-94439-1-P1.

Gerjuoy, E., and Stein, S. (1955).—Westinghouse Research Laboratories Sci. Res. Pap. 60-94439-1-P2.

Hall, Barbara I. H. (1955).—*Aust. J. Phys.* **8**: 468.

Massey, H. S. W., and Burhop, E. H. S. (1952).—"Electronic and Ionic Impact Phenomena." p. 279. (Clarendon Press: Oxford.)

THE SCATTERING OF 121 keV ELECTRONS AND POSITRONS

By L. J. Tassie*

[*Manuscript received November 23, 1955*]

Summary

The angular distribution of the single scattering of 121 keV positrons in gold and of 121 keV electrons and positrons in argon is calculated for an exponentially screened field. The single scattering distributions are integrated numerically to give mean square angles of multiple scattering, and hence the percentage difference in the r.m.s. angles of multiple scattering of electrons and positrons.

I. Introduction

Calculations have previously been carried out by Mohr and Tassie (1954*a*)† of the effect of screening on the elastic single scattering of electrons by gold, and these results were applied to the problem of multiple scattering (Mohr and Tassie 1954*b*)‡. In the latter work, the single scattering of 121 keV positrons by gold was also calculated making approximate estimates for the phase shifts. This calculation of positron scattering is repeated using more accurate phase shifts. The single scattering distribution is then integrated numerically to give the mean square angle of multiple scattering, and the difference in the multiple scattering of electrons and positrons obtained by comparison with the results for electron scattering (paper B).

As the available experimental results are for light elements, a similar calculation is carried out for scattering by argon. Although experimental results are for somewhat higher energies, 150 keV upwards, the calculation is made for incident particles of energy 121 keV, the same energy as in the calculation for gold. Any effects of screening by the atomic electrons would be expected to be smaller for higher energies, and any effects of the electron spin to be larger. The Hartree field of argon (Hartree and Hartree 1938) was approximated by a one-term field (5). Although a one-term field is a good approximation to the Hartree field of gold, in the case of argon additional terms of the form $br\exp(-\beta r)$ are really necessary (Ibers and Hoerni 1954). However, the use of a one-term field greatly simplifies the calculation and is sufficient to determine the nature of the effects of screening on the scattering.

II. General Theory

The differential cross section for the scattering by a potential $V(r)$ is given by (Mott and Massey 1949)

$$I(\theta)=|f|^2+|g|^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

---

* Physics Department, University of Melbourne.

† Hereafter referred to as paper A.

‡ Hereafter referred to as paper B.

where

$$2ikf(\theta)=\Sigma[(l+1)\{\exp(2i\eta_l)-1\}+l\{\exp(2i\eta_{-l-1})-1\}]P_l(\cos\theta),$$
$$2ikg(\theta)=\Sigma\{\exp(2i\eta_{-l-1})-\exp(2i\eta_l)\}P_l^1(\cos\theta), \quad \ldots\ldots\ldots\ldots\ldots \quad (2)$$

$k^2=4\pi^2(W^2-m_0{}^2c^4)/h^2c^2$, W being the total energy. η_l is the phase shift such that $\sin(kr-\frac{1}{2}l\pi+\eta_l)$ is the asymptotic form of that solution G_l of the equation

$$G_l''+\{k^2-l(l+1)r^{-2}-U_l(r)\}G_l=0, \quad \ldots\ldots\ldots\ldots \quad (3)$$

which vanishes at the origin.

U_l is the modified Dirac potential

$$U_l(r)=8\pi^2WV/h^2c^2-4\pi^2V^2/h^2c^2-(l+1)\alpha'/\alpha r+3\alpha'^2/4\alpha^2-\alpha''/2\alpha, \quad \ldots \quad (4)$$

where $\alpha=2\pi(W-V+m_0c^2)/hc$.

An exponentially screened field was used for the atomic field V

$$V(r)=Z_p\varepsilon^2/r,$$
$$Z_p=Z\exp(-\chi r/a_0), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

where $a_0=h^2/4\pi^2m_0\varepsilon^2$. The same field was used for gold as in paper A, namely, (5), with $\chi=3$. The Hartree field of argon (Hartree and Hartree 1938) was fitted approximately by (5) with $\chi=2$. A comparison of this field with the Hartree field and with the Thomas-Fermi field as used by Molière (1947) is given in Figure 1.

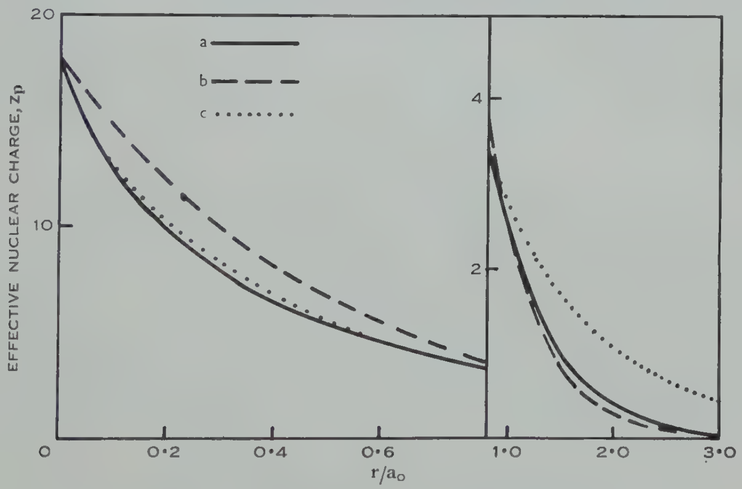


Fig. 1.—Variation with distance of the effective nuclear charge Z_p for argon. *a*, Hartree field; *b*, one-term field, equation (5); *c*, Molière's three-term field approximating to the Thomas-Fermi field.

III. Determination of the Phase Shifts

The same procedure was used for computing the phase shifts as in paper A. The phase shifts η_l^0 for only the first term in the effective potential (4) (i.e. the Schroedinger phase shifts) were first obtained and then corrected for the effects of spin as given by further terms in the effective potential. The region of validity of the approximations used has been previously established by comparison with numerical integration for the scattering of electrons by gold, and checks by numerical integration and by the WKB method were carried out in only a few cases. The K_0 approximation,

$$\eta_l^0=(\gamma Z/ka_0)K_0\{\chi(l+\tfrac{1}{2})/ka_0\}, \quad \ldots\ldots\ldots\ldots\ldots \quad (6)$$

where

$$\gamma=(1-v^2/c^2)^{-1}$$

was used with the correction due to Pais (1946) for $l \geqslant 8$. For $l<8$, the η_l^0 were obtained for the scattering of electrons by argon by comparison with the phases for a Coulomb field and the difference in the $l=1$ phase for electrons and positrons, 0·006, was determined by the WKB method. For the scattering of positrons by gold, the zero-order phase was obtained by numerical integration and the other phases by interpolation, a procedure which was checked by calculating η_5^0 by the WKB method.

The effect of spin on the phases, $\eta_l-\eta_l^0$, was obtained from the effect of spin on the Coulomb phases, corrected for screening for argon by an approximate analytical correction (Appendix I). For gold this spin effect was obtained by the WKB method for $l=0, 5$. The modification of the spin effect by screening is small, e.g. the modification is largest for $l=0$ for gold where $\eta_0-\eta_0^0=0\cdot181$ for the Coulomb field, and 0·148 for the screened field, and any error introduced by the use of the above procedure is therefore small.

IV. Angular Distribution of Single Scattering

The Mott series (**2**) was summed for large angles ($\theta \geqslant 30°$) using the same procedure as in paper A, and for small angles using method (*a*) of paper B. For the small angle scattering of positrons by gold, the difference in the series (**2**) for electrons and positrons was summed, and the results of paper A for the summation for electrons were used.

The results obtained for the scattered intensity are shown in Figures 2 and 3 as the ratio R to the relativistic Rutherford scattering $Z^2\varepsilon^4 \operatorname{cosec}^4\frac{1}{2}\theta/4\gamma^2m_0^2v^4$. The values of R for the Coulomb field of mercury are taken from Massey (1942). For light elements Curr (1955) gives R as a power series in $\alpha=Z/137$ and $q=\alpha/3$. For the scattering by argon at 121 keV, $\alpha=0\cdot13$ and $q=0\cdot22$, and the power series converges rapidly.

For large angles the scattering by the screened field is the same within the accuracy of the calculation as the scattering by a Coulomb field. The ratio of the scattered intensities for electrons and positrons is approximately the same for the screened and Coulomb fields of argon. However, for gold this ratio is larger for scattering by the screened field than by the Coulomb field.

The results for the scattering through small angles are compared in Figure 3 with the results of paper B for the scattering of 121 keV electrons by gold and with the Molière single scattering distributions. Molière (1947) gives two single scattering distributions. One is obtained from an empirical interpolation formula which interpolates between the Born approximation and the classical approximation. This is labelled M' in Figure 3. The other is a more approximate distribution of simpler form,

$$R=\theta^4/(\theta^2_{\text{min.}}+\theta^2)^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

where $\theta_{\text{min.}}$ is determined by comparison with his more accurate distribution. This is labelled M in Figure 3.

For argon, the scattering distributions given by the Born approximation for the one-term field, for the Hartree field (Hartree and Hartree 1938), and for the Thomas-Fermi field in the form used by Molière (1947) are also shown. This indicates that the difference between the results calculated here and the Molière scattering distributions is largely due to the difference in the fields used. The one-term field is a rather poor approximation to the atomic field of argon and the Thomas-Fermi field used by Molière is more appropriate for $\theta > 1°$.

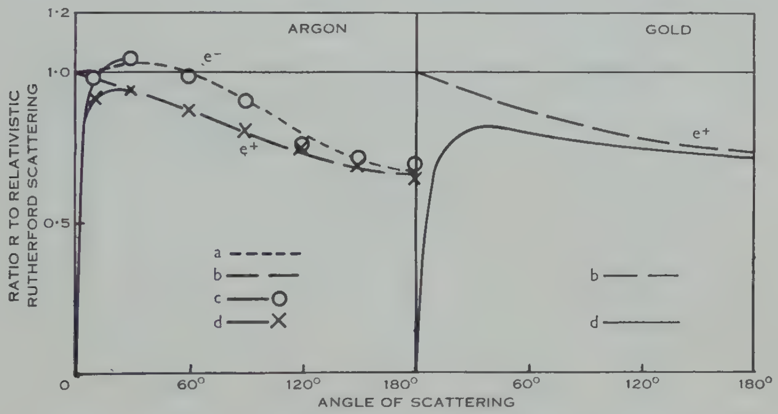


Fig. 2.—Angular distribution of single scattering of 121 keV electrons and positrons. *a*, Electrons, Coulomb field; *b*, positrons, Coulomb field; *c*, electrons, screened field; *d*, positrons, screened field.

V. Multiple Scattering

The mean square angle of multiple scattering is given by

$$\theta^2_{\text{r.m.s.}}=\overline{\theta^2}=\int_0^{\theta_{\text{max.}}}\theta^2P(\theta)\mathrm{d}\theta, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

where $P(\theta)=2\pi NtI(\theta)\sin\theta$; N is the number of atoms per unit volume and t is the thickness of scattering material.

Values of $\theta_{\text{r.m.s.}}$ were obtained by numerical integration of (8) using the single scattering distributions for electrons and positrons. This allows the determination of the difference in $\theta_{\text{r.m.s.}}$ for positrons and electrons as a function of $\theta_{\text{r.m.s.}}$ for electrons. The results for this are shown in Figure 4 as $100(\theta^-_{\text{r.m.s.}}-\theta^+_{\text{r.m.s.}})/\theta^-_{\text{r.m.s.}}$. As would be expected, this difference is greater for gold than for argon.

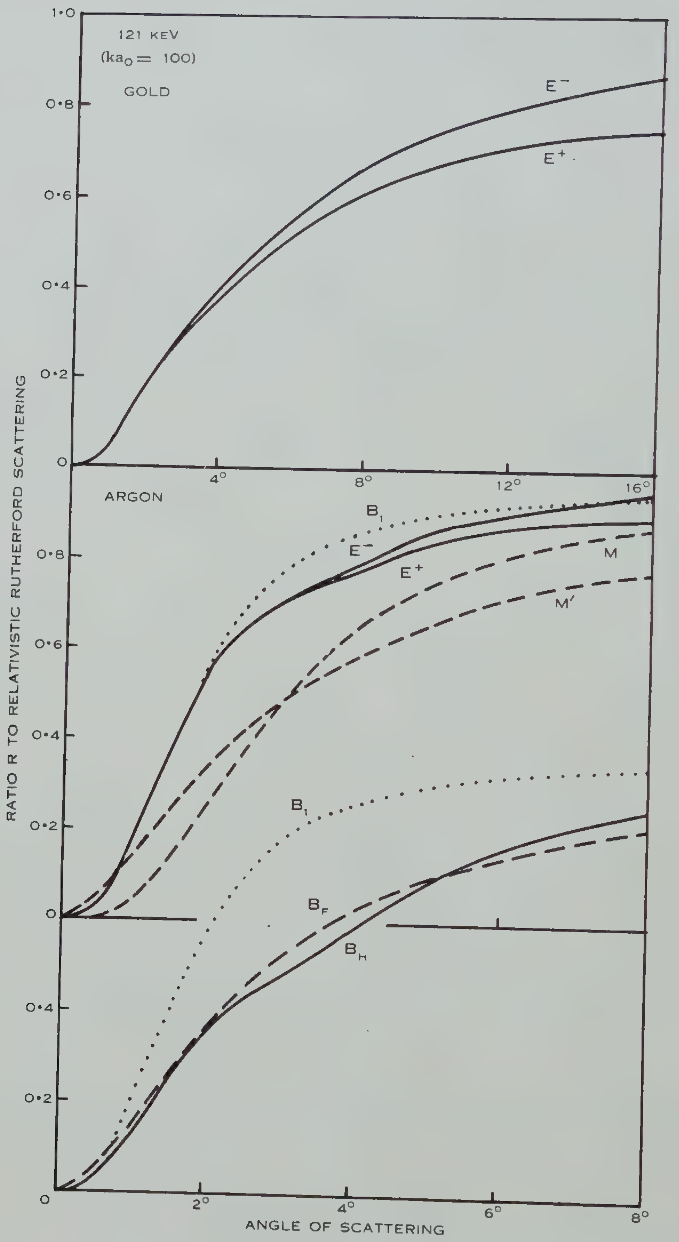


Fig. 3.—Angular distribution of single scattering of 121 keV electrons and positrons. B_1, Born approximation with one-term field, equation (**5**); B_H, Born approximation with Hartree field; B_F, Born approximation with Thomas-Fermi field. *M*, Calculated from the simplified Molière theory for application to multiple scattering, equation (**7**) (Molière 1947, equations (**9.1**) and (**9.3**)); *M′*, calculated from Molière's more accurate theory (Molière 1947, equation (**8.6**)). E^-, E^+, calculated using the exact theory with the one-term field for electrons and positrons respectively; E^- for gold is taken from Mohr and Tassie (1954*b*).

The results obtained here are in qualitative agreement with previous calculations of paper B and of Mohr (1954), in predicting that the percentage difference in $\theta^{+}_{\text{r.m.s.}}$ and $\theta^{-}_{\text{r.m.s.}}$ is small. The latter calculation using the second Born approximation gives for an energy of 1 MeV and $\theta_{\text{r.m.s.}}=20°$ a difference of $1 \cdot 7$ per cent. for $Z/137=0\cdot15$ and 5 per cent. for $Z/137=0\cdot58$, a difference which increases as the energy decreases. This would seem to indicate that $(\theta^{-}_{\text{r.m.s.}}-\theta^{+}_{\text{r.m.s.}})/\theta_{\text{r.m.s.}}$ has a maximum value somewhere between 1 MeV and 121 keV. However, Mohr uses the theory of Molière (1948) to obtain a mean value of $\theta_{\text{r.m.s.}}=(\theta^{-}_{\text{r.m.s.}}+\theta^{+}_{\text{r.m.s.}})/2$, a procedure which cannot safely be used for the present case.

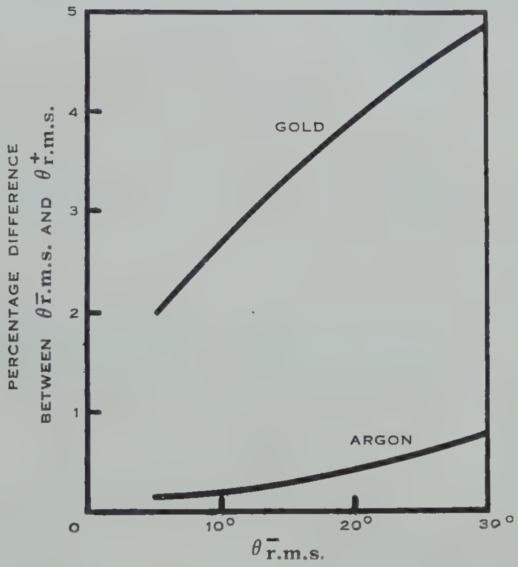


Fig. 4.—The percentage difference between the root mean square angle of multiple scattering of 121 keV electrons and positrons as a function of the r.m.s. angle for electrons, for scattering by gold and by argon.

The small values of $(\theta^{-}_{\text{r.m.s.}}-\theta^{+}_{\text{r.m.s.}})/\theta_{\text{r.m.s.}}$ predicted by theory are in disagreement with the experimental results of Groetzinger, Humphrey, and Ribe (1951), who report a difference of about 10 per cent. in $\theta^{-}_{\text{r.m.s.}}$ and $\theta^{+}_{\text{r.m.s.}}$ for argon. However, the later and more accurate experiments of Cusack and Stott (1955) indicate no significant difference in the multiple scattering of electrons and positrons in argon.

VI. Acknowledgments

The author would like to express his thanks to Associate Professor C. B. O. Mohr for his interest and encouragement in this work. He is indebted also to the Commonwealth Scientific and Industrial Research Organization for an Australian Studentship.

VII. References

Curr, R. M. (1955).—*Proc. Phys. Soc. Lond.* A **68** : 156.

Cusack, N., and Stott, P. (1955).—*Phil. Mag.* **46** : 632.

Groetzinger, G., Humphrey, W., Jr., and Ribe, F. L. (1951).—*Phys. Rev.* **85** : 78.

Hartree, D. R., and Hartree, W. (1938).—*Proc. Roy. Soc.* A **166** : 450.

Ibers, J. A., and Hoerni, J. A. (1954).—*Acta Cryst.* **7** : 405.

Massey, H. S. W. (1942).—*Proc. Roy. Soc.* A **181** : 14.

Mohr, C. B. O. (1954).—*Proc. Phys. Soc. Lond.* A **67** : 730.

Mohr, C. B. O., and Tassie, L. J. (1954*a*).—*Proc. Phys. Soc. Lond.* A **67** : 711.

Mohr, C. B. O., and Tassie, L. J. (1954*b*).—*Aust. J. Phys.* **7** : 217.

Molière, G. (1947).—*Z. Naturf.* **2a** : 133.

Molière, G. (1948).—*Z. Naturf.* **3a** : 78.

Mott, N. F., and Massey, H. S. W. (1949).—" The Theory of Atomic Collisions." 2nd Ed. p. 76. (Oxford Univ. Press.)

Pais, A. (1946).—*Proc. Camb. Phil. Soc.* **42** : 45.

Appendix I

Effect of Spin on the Phase Shifts

We require

$$S_l=\eta_l-\eta_l^0,\quad S_{-l-1}=\eta_{-l-1}-\eta_l^0$$

for a field of the form (5). This can be obtained by applying the WKB method to the modified Dirac potential U_l, equation (4), and assuming U_l small compared with $k^2a_0^2$. The result of this procedure is then corrected to give the exact Coulomb field values of S as $\chi\to 0$, giving the final result :

$$S_l=P_1(l)C_l^{(1)}+P_2(l)C_l^{(2)}+C_l^{(3)},$$
$$S_{-l-1}=P_1(l)C_{-l-1}^{(1)}+P_2(l)C_{-l-1}^{(2)}+C_{-l-1}^{(3)},$$

where

$$C_{-l-1}^{(1)}=\tfrac{1}{2}\pi(l-\rho_l),$$
$$C_{-l-1}^{(3)}=\arg\Gamma(\rho_l+1-iq)-\arg\Gamma\,(l+1-iq),$$

($C_l^{(1)}$ and $C_l^{(3)}$ are obtained by replacing l by $l+1$) ;

$$-C_l^{(2)}=\arctan\{q/(l+1)\}-\tfrac{1}{2}\arctan\,(q/\rho_{-l-1})-\tfrac{1}{2}\arctan\{q'/(l+1)\},$$
$$C_{-l-1}^{(2)}=\tfrac{1}{2}\arctan\,(q/\rho_l)-\tfrac{1}{2}\arctan\,(q'/l),$$
$$\rho_l=(l^2-\alpha^2)^{\frac{1}{2}},\ \alpha=Z/137,\ q=\alpha/\beta,\ q'=(1-\beta^2)^{\frac{1}{2}}\ ;$$
$$P_1^{(l)}=\{1-2\pi^{-1}(-2b\ln 2b+2\cdot 2320b+8b^3/3)\},$$
$$P_2^{(l)}=\{1+b^2\ln b-0\cdot 3508b^2-\tfrac{1}{4}b^4\},$$

where $b=\chi(l+\frac{1}{2})/ka_0$.

18·3 Mc/s RADIATION FROM JUPITER

By C. A. Shain*

[*Manuscript received August* 23, 1955]

Summary

Burke and Franklin's discovery of radio emission from Jupiter has been confirmed. Examination of old records has shown that in 1950–51 the radiation came in groups of bursts of very high intensity. Bursts have durations of the order of a minute or less; groups, of an hour. Because of the remarkably close relation between active periods and the period of rotation of Jupiter, it is inferred that the source at the time was very localized. Its identification with a visual disturbance in the South Temperate Belt is very probable.

It is pointed out that occultation by Jupiter's satellites may help to locate the sources of radiation, both in position and in height relative to the visible surface, and that Jupiter radiation should be a valuable tool for studying the outer regions of the solar corona.

I. Introduction

During the rapid development of radio astronomy in the last 10 years, the question has sometimes been raised as to whether radiation from any of the planets could be detected. The detection of thermal radiation would appear to be at present impracticable, but it has been suggested (Higgs 1951) that electrical discharges analogous to terrestrial lightning flashes may occur in the atmosphere of Venus and that radiation from such discharges may be detectable. In 1955, however, came the quite unexpected announcement by Burke and Franklin (1955) that very intense radiation at 22 Mc/s had been received from a direction in the sky which was close to the position of Jupiter, the Right Ascension of the source changing with that of the planet.

Following this remarkable discovery of radio radiation from Jupiter, observations were begun in this Laboratory and at the same time records of cosmic noise which had been made during the past few years were searched for signs of Jupiter radiation. Although there was no trace found of any radiation from Jupiter on records at higher frequencies, some records taken at 18·3 Mc/s in 1950–51 showed series of bursts which had previously been passed over as terrestrial interference. Detailed analysis has now shown these to be radiation from Jupiter, thus confirming Burke and Franklin's discovery. Study of the same observations has added an important new fact; for a period of at least one and a half months in 1951 the radiation came from a very localized region on Jupiter.

The new observations are few and will not be described here in detail; they too show Jupiter radiation, though less frequently than the earlier records.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

The present paper discusses the older observations. After a general description of the records and of the equipment used, two series of records are considered in detail. These give evidence that the radiation came from a small region of the planet. A probable identification of one such region is put forward. Finally, the usefulness of Jupiter radiation for the study of radio propagation conditions near the Sun is pointed out.

II. Observations

(a) *Circumstances*

The records discussed in this paper were made in the course of a study of cosmic noise at 18·3 Mc/s ; the equipment used has been described previously (Shain and Higgins 1953). The aerial consisted of an array of 30 half-wave dipoles and the direction of maximum sensitivity in the meridian plane could be changed. The beamwidth to half-power was 17°. The feeders were arranged so that two receivers could be connected to the aerial, one using the aerial with the maximum response in the meridian plane, the other with zero response in the meridian. When suitable records were available from both receivers, a comparison of the signal powers received in the two receivers could be used to find the direction of the source of the signal.

With this system records were taken almost daily for about a year up to June 1951. The aerial direction was changed at frequent but irregular intervals in the course of the cosmic noise programme, and a watch was kept for peculiarities on the records, especially for bursts of solar noise and for variations caused by abnormal ionospheric attenuation. Quite frequently there was interference from atmospherics and radio stations ; thus no particular significance was attached to the occurrence of occasional groups of bursts during the night. However, after the announcement by Burke and Franklin, a review of these records showed that on a number of occasions groups of large bursts occurred when Jupiter was in the aerial beam. In the periods October 16–November 30, 1950 and February 9–April 18, 1951, there were 64 days on which records were suitable for the detection of Jupiter radiation (correct aerial direction, no obvious interference, etc.).*

Study of these records identified the source of the radiation as Jupiter and facilitated recognition of Jupiter radiation in a second series of records, from August 15 to October 2, 1951, taken under somewhat different conditions. In June 1951 part of the original aerial had been dismantled, but a single receiver was used to measure the noise picked up by the remaining 10 dipoles, which had an aerial diagram narrow in the north-south direction but very broad in the east-west direction. These records, which permitted the observation of Jupiter for almost a whole rotation on any one day, have provided information which has led to the conclusion that the radiation originated in an area of small extent on the surface of Jupiter.

The two series of observations will be discussed in turn.

* All dates and times in this paper are given in 150° E. time unless otherwise stated.

(*b*) *First Series—Identification of the Source with Jupiter*

Figure 1 shows six selected records, for which the aerial had maximum sensitivity in the meridian, so arranged that for each the time of transit of Jupiter lies on the line joining the two arrows. It will be seen that the records are smooth except for the short interval of about 2 hours during which Jupiter was passing through the aerial beam. It is apparent that over several months the time of maximum noise changed from late evening to midday. The changing time interval between the groups of bursts and the passage of part of the Milky Way through the aerial beam (on the right of the records) can also be seen, indicating that the Right Ascension of the source is changing.

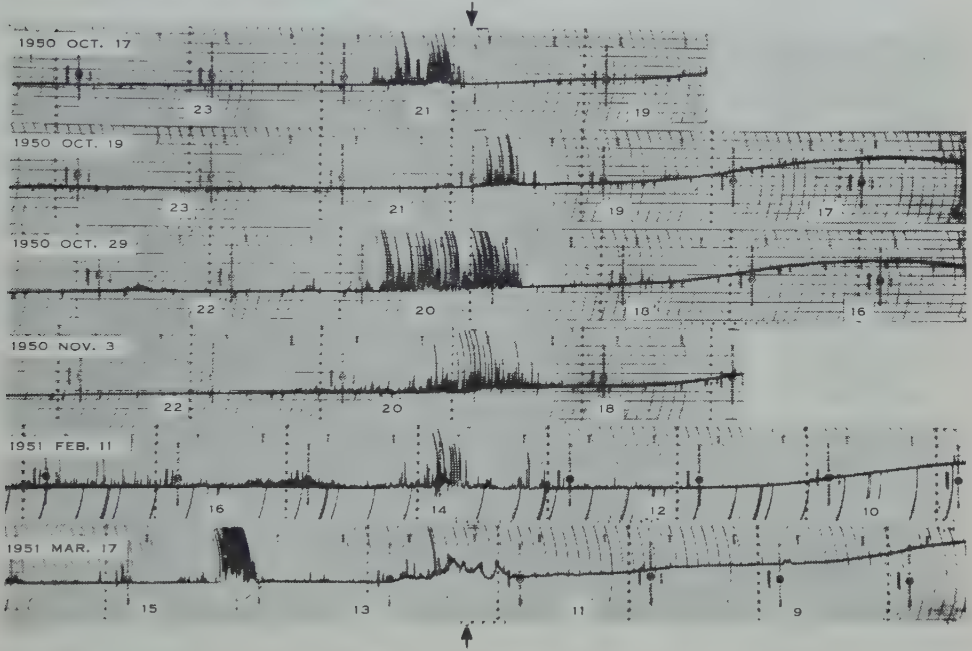


Fig. 1.—Records showing strong bursts near the time of transit of Jupiter. Dates and times are given on each record. The records have been arranged so that the times of transit of Jupiter are on a line joining the arrows. Atmospheric noise is present on February 11, 1951, from 15^h onwards, and radio station interference on March 17, 1951, between 14^h and 15^h.

Most of the records show violent fluctuations, often going off-scale at an intensity greater than 5×10^{-21} W m^{-2} (c/s)$^{-1}$, but the appearance changed for the later records, which were taken when the time of transit of the source was near midday. This change of appearance will be referred to again later. The records were made by a meter having a time constant of the order of 1 sec; thus transients of shorter duration, if present in the radiation, would not be recorded. Notes written on some records indicate that many bursts sounded like " swishes " (similar to solar noise bursts, with which they were confused) and therefore the recorder followed these noise variations faithfully. On some other occasions there were sounds suggesting the presence of short impulses which the recorder would not follow.

As mentioned above, there were 64 records which were suitable for the detection of bursts near the time of transit of Jupiter. On 17 of these there were intense bursts having characteristics as shown in Figure 1, and on a further 17 there were definite but weaker bursts. On the remaining days no sign of such fluctuations could be detected. This appearance of bursts on only about half the days agrees with another deduction from Figure 1, namely, that on those days on which bursts were received they were not generally present on the records for the full time that the system was sensitive to radiation from a source moving through the aerial beam.

Confirmation of the tentative conclusion that the records were of radiation which had originated on Jupiter came from the detailed comparison of the records from the two receivers, which gave fairly accurate positions of the source of the noise on a particular day. Two pairs of records are shown in Figure 2. For the upper record in each case the aerial had maximum sensitivity in the meridian, while for the lower records there was a zero in the meridian. In Figure 2 (*b*) the bursts commenced before the source had reached the central plane of the aerial and the passage of the source across this central plane can be seen by the absence of bursts below the arrow on the lower record. In Figure 2 (*a*) the bursts did not begin until after the source had passed the central plane of the aerial and the time of zero response cannot be read off directly from the record. However, by finding the ratio of the amplitudes of individual bursts on the two records and comparing these with the ratio to be expected at the same times by calculation from the known aerial diagrams, the time of passage of the source across the central plane of the aerial could be deduced with an accuracy of better than ± 4 min (an angular accuracy of $\pm 1°$). This direction-finding method was checked by application to cosmic noise sources and the Sun. It was not always possible to use this technique as it required good records, free of interference, from both receivers, but on the 13 days for which there were suitable records the origin of the bursts was located within 1° of Jupiter. In this way some solar bursts could be distinguished from Jupiter radiation ; the few bursts of short duration on the record for March 17 (Fig. 1) were found to be from the Sun whilst the large slow variations came from Jupiter.

Some correlation was found between the times of appearance of the noise and the appearance on the visible disk of Jupiter of certain regions of longitude. When all the records were taken together, there was a greater chance of receiving the noise when longitude 110° (System II)* was the central meridian on Jupiter's disk. The correlation was not strong enough for detailed study, but it did show that the Red Spot was not a prominent source of radio noise.

---

* The equatorial regions of Jupiter rotate faster than the remainder of the planet and within each region there are slight variations in period. Two conventional systems of longitude are therefore used. System I is applicable to the equatorial regions and System II to the remainder of the planet. The adopted daily motions and rotation periods of the zero meridians of the two systems are :

System I	.. 877°·90	$9^{h}\ 50^{m}\ 30^{s}\cdot 003$
System II	.. 870°·27	$9^{h}\ 55^{m}\ 40^{s}\cdot 632$

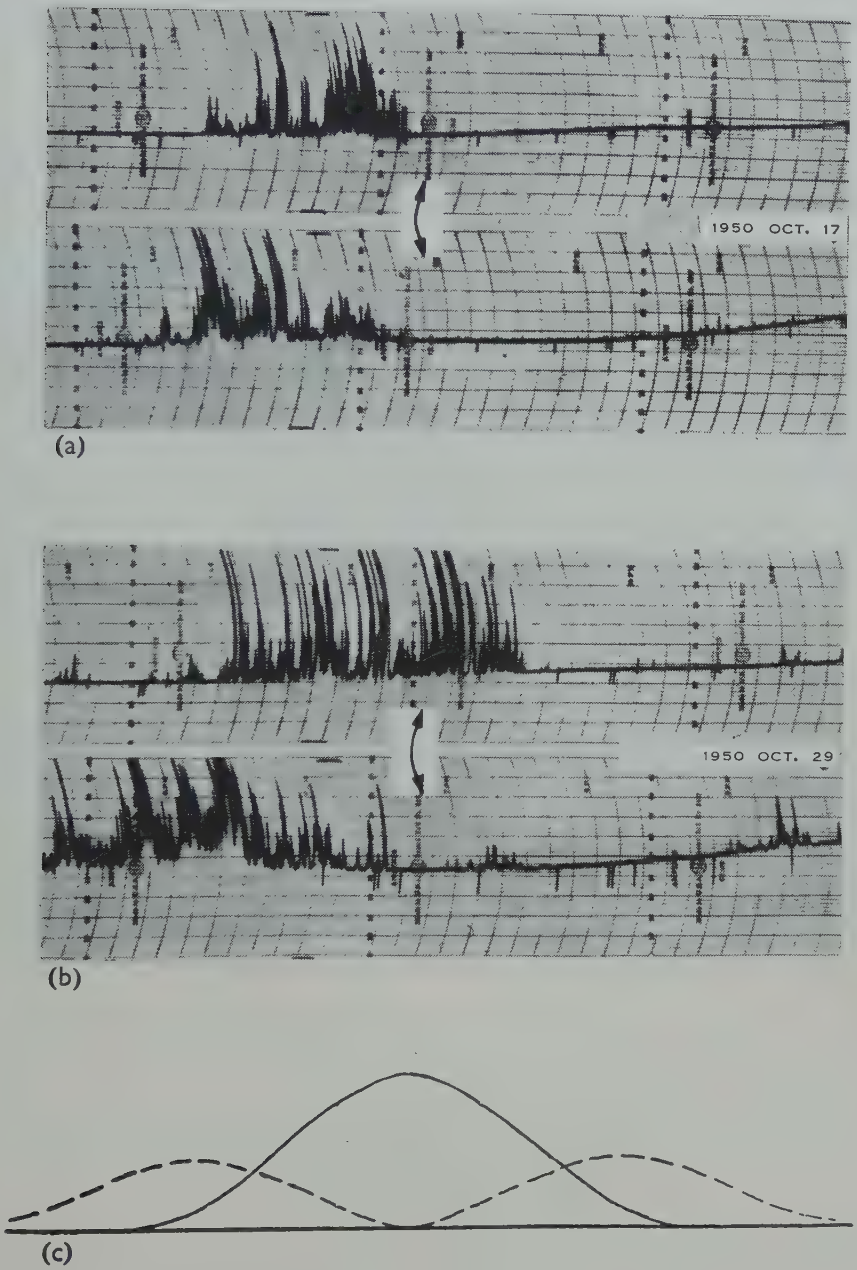


Fig. 2.—Radiation from Jupiter, recorded with two aerials simultaneously, on (*a*) October 17, 1950 and (*b*) October 29, 1950. (*c*) shows the aerial sensitivity. In each case the upper record was made using an aerial with maximum sensitivity in the meridian plane (full curve in (*c*)), while for the lower records there was a zero in the meridian plane (dashed curve in (*c*)). Time of transit of Jupiter is indicated by the arrows.

Having established the characteristics of the radiation from Jupiter, one could recognize with reasonable certainty the same radiation on the records of the second series, made with the broader aerial beam.

(c) *Second Series—August 15 to October 2, 1951*

(i) *Description of the Records.*—After modification in June 1951 the aerial was capable of receiving signals from an extraterrestrial source for nearly 8 hours per day. The beam was still narrow in the north-south direction so that there was discrimination between sources at different declinations, but there were no facilities for accurate direction-finding in the east-west direction.

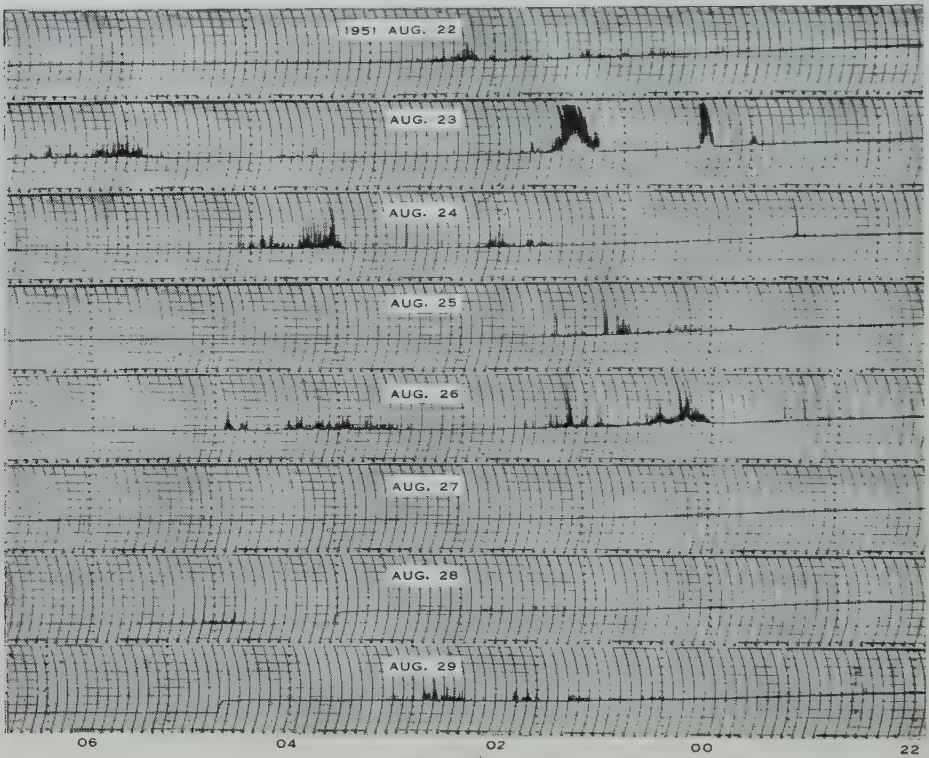


Fig. 3.—Records taken with a very broad aerial beam on successive days (the dates refer to the times after midnight). The time of transit of Jupiter is roughly in the centre of each record.

There is station interference on the record for August 23 at about 00h and 01h, and also probably from about 00h to 00h 30m on August 26. All the other groups of bursts are believed to be from Jupiter. During the later part of the records for August 28 and 29 the receiver sensitivity was low.

During the period under consideration the aerial was sensitive to noise from a source on Jupiter's declination and examination of the records showed that again there were groups of bursts near the time of transit of Jupiter. A series of records is shown in Figure 3. These groups of bursts were similar in appearance to those on the earlier records which had been shown to come from

Jupiter, and our experience has shown it to be most unlikely that terrestrial interference would be recorded so regularly at midnight during winter and spring. There were two direct checks that these groups of bursts were actually from Jupiter. Firstly, although the broad aerial beam made useful direction-finding impossible on any one day, the time of transit of Jupiter changed by nearly 4 hours between the middle of August and the beginning of October, and the time of occurrence of the bursts did the same. Secondly, as will be shown later, there was a very close relation between the occurrence of these bursts and the rotations of Jupiter.

With the broad aerial beam, Jupiter was in the beam for a time which permitted almost a complete rotation of the planet. Even if only a fraction of the planet were radiating, it would be expected that noise would be received on a greater proportion of days than was found with the narrower beamed aerial.

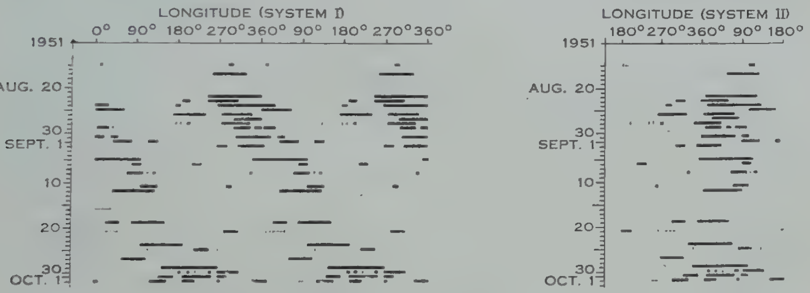


Fig. 4.—Periods of occurrence of 18·3 Mc/s radiation from Jupiter plotted against longitude of the central meridian at the time of observation.

This was found to be the case, noise being detected on 27 out of the 30 days on which there were suitable records. Figure 3 clearly shows that the groups of noise bursts were generally only one or two hours long. Since this is only a small fraction of the rotation period of Jupiter, there must be hour-to-hour variations in the intensity of the noise radiated by the planet.

(ii) *The Relation between the Times of Occurrence of Bursts and the Rotation of Jupiter.*—For each record the times of activity were noted and charts, shown in Figure 4, were drawn up to indicate any correlation with the rotation of Jupiter. For each day a thick line has been drawn under the longitude of the central meridian during the times when bursts were observed. It will be seen that these lines are almost directly under one another when plotted against System II longitudes and show a steady drift towards increasing longitudes in System I. Closer examination shows that there is actually a small negative drift in the System II diagram. These drifts imply that the source of radiation had a rotation period slightly shorter than that adopted for the calculation of System II longitudes ; the period of rotation is found to be $9^h\ 55^m\ 13^s$ with an estimated probable error of $\pm 5^s$.

Allowing for the slight drift in longitudes, all the lines in Figure 4 were superimposed to give a histogram of the frequency of occurrence of the noise for 5° intervals of central meridian longitude. This is shown in Figure 5, the longitudes being System II longitudes on August 14. It is seen that for a band of longitudes centred on 67° and extending from about 0 to 135° the frequency of occurrence was much greater than outside this band. It should be pointed out that there were about 120 rotations of Jupiter during the period of the observations on which this diagram is based, so that the probability that the effect observed was due to chance is extremely small.

The observations for the period February to April 1951 also indicated the same rotation period, within the experimental uncertainty, but the actual longitudes concerned were about 80° smaller than would have been obtained by extrapolation backwards from the second series observations.

III. Discussion of the Observations

(a) *Size of the Source*

If the radio noise came from a single source radiating in all directions, one would expect that the histogram in Figure 5 would have a flat maximum, cutting off sharply at longitudes 180° apart. Figure 5 does show a fairly flat maximum, but the boundaries are separated by only about 135°. There are several possible reasons for this restriction of the angle during which radiation may be received at the Earth. One possibility is that the emission of radio bursts is stimulated by solar visible (more probably ultraviolet) radiation. Viewed from Jupiter, the Earth is always close to the Sun so that the Sun rises at any point on Jupiter at very nearly the same time as the Earth comes into view. Therefore, if there were any delay after sunrise before the source radiated strongly, this would be shown as a restriction in the range of central meridian longitudes for which the noise was received.

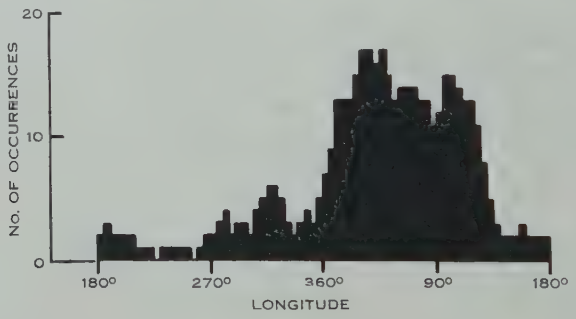


Fig. 5.—Frequency of occurrence of 18·3 Mc/s radiation from Jupiter for intervals of 5° in longitude (based on a rotation period of $9^h 55^m 13^s$). The longitudes are System II longitudes on August 14, 1951.

Depending on the height of the source of the radiation, atmospheric refraction may be more important on Jupiter than on the Earth. Refraction would tend to extend the time during which the noise source would be " visible ", but this effect may be counteracted by the action of the atmosphere as a diverging lens, decreasing the intensity when the source is near the edge of the disk. A further possibility is that an ionosphere on Jupiter would cut off radio radiation when the Earth was at low altitudes as seen from the source.

Alternatively, if the source were actually in the ionosphere and near the level of maximum ionization density, refraction would operate in the correct direction. Jupiter probably has an ionosphere, but at present its characteristics are quite unknown.

In any case the emitting region was probably small in extent to give an "emission polar diagram" of much less than 180°. It therefore appeared worth while to attempt to identify the source with visual features of Jupiter.

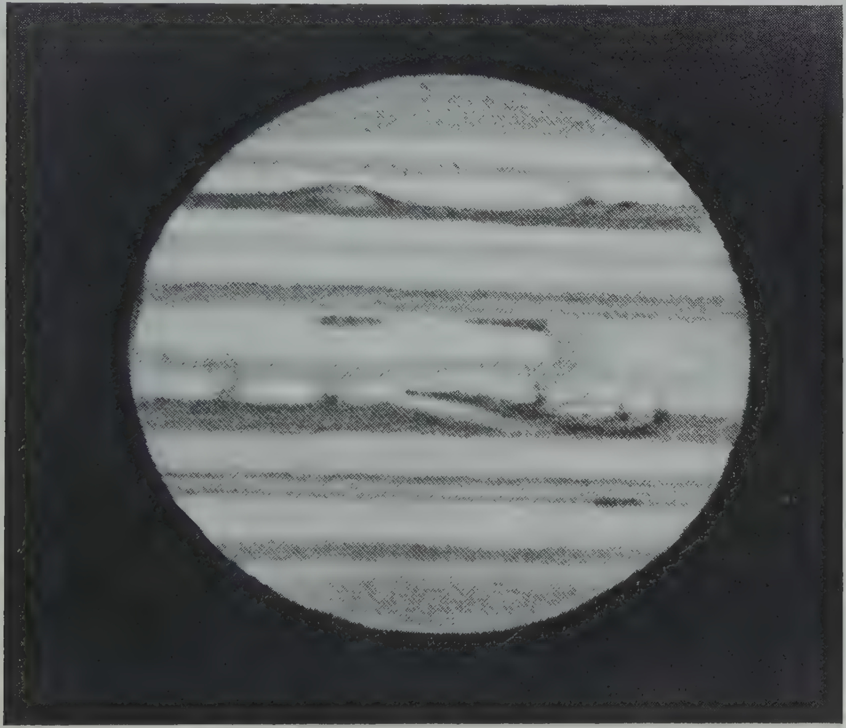


Fig. 6.—Jupiter on September 22, 1951 at 02^h U.T. The longitudes on the central meridian are 161° (System I), and 41° (System II). South is at the top.

(b) *Identification of the Radio Emitting Region*

The basic data were that the longitude of the source on August 14 was 67° (System II) and the rotation period was $9^h\ 55^m\ 13^s$, the region being still active when observations ceased on October 2. The rotation period is important; it should be remembered that the two systems of longitude are fixed by convention and actual visible markings commonly drift at differing rates.

A drawing of Jupiter by du Martheray (du Martheray and Antonini 1952) is shown in Figure 6. Fox (1952) has given a summary of contemporary observations of Jupiter which were communicated to the Jupiter Section of the British Astronomical Association, and from this it appears very likely that the source of the radiation was located in the South Temperate Belt (the dark belt near the top of the drawing in Figure 6). A group of spots in this belt, which were observed for several months, and which were considered to be of a long-enduring character, had an observed rotation period of $9^h\ 55^m\ 13^s$, just the same as that of the radio source. All the other regions which move with System II were either faint with no certain markings or moved slightly slower than System II. The most prominent belt, with many markings, was the North Equatorial Belt, but this rotates with System I. Fox gave a sketch by E. J. Reese (reproduced in Fig. 7) which shows the most visually active region of the South Temperate Belt at the end of November 1951. The radio observations ceased on October 2, but, allowing for the continual drift in longitude, the longitude of the radio source on November 30 would have been 354° (System II); this corresponds with the longitude of the white spot labelled DE in Figure 7. This spot appears in Figure 6 as the indentation near the central meridian in the southern edge of the South Temperate Belt. Also, reference to Figure 5 shows that there was some radio noise from sources at smaller longitudes than the main source. The positions of these minor sources would have extended to longitude 260° (System II) by November 30. Therefore, although the identification is not proved beyond doubt, it seems very probable that the visually disturbed region in the South Temperate Belt was responsible for the radio radiation, the most intense source lying in one of the white spots in the southern edge of the belt.

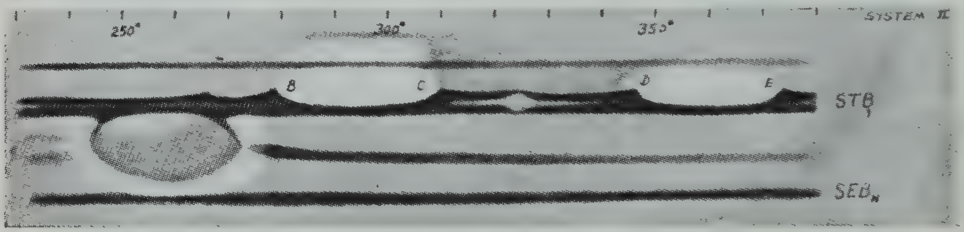


Fig. 7.—A drawing of part of Jupiter's disk on November 29–30, 1951 U.T. An extrapolation of the position of the radio source to November 30 would give its longitude as 354°.

(c) *Occultation by Satellites*

During the period for which observations were available there were three passages of Galilean satellites across the meridian containing the radio source. The records at these times were examined to see whether there were any marked changes, corresponding to occultations, which would help to locate the source. On one occasion there was no effect at all, and on a second occasion there was only a very doubtful suggestion of an occultation. In both cases the position of the satellite was always far from the position of the source deduced in the preceding section.

On the third occasion, during the transit of Satellite II on September 24, there was a sudden cessation of the noise as the satellite reached the meridian of the source, at $03^h\ 10^m$. This may have been a coincidence, since the noise had already been received for over 2 hr, but on one or two other days noise had been received continuously for over 4 hr. At $03^h\ 10^m$ the source was still within 25° of the central meridian and as there was weak noise again 32^m later, there is a reasonable probability that the sudden decrease in intensity was due to an occultation of the source. The portion of the visible disk covered by the satellite was almost exactly the probable position of the source deduced previously.

If an occultation did occur, its duration, for a source of small extent, should have been only 16^m; in fact there was no noise for about half an hour. The obvious explanation of this discrepancy would be that slow fading in the intensity of the radiation masked the actual end of the occultation. An interesting alternative possibility is that the source was located well above the visible surface of the planet. The duration of the occultation would be extended to 32^m if the source were at a height of $0 \cdot 1R$, R being the radius of the visible surface; that is, at a height of about 7000 km. Also, there was no effect during the passage across the source of the shadow of the satellite, which occurred at about $01^h\ 20^m$. If this result were genuine it would rule out direct solar stimulation of the source of the radiation.

(*d*) *Propagation Conditions between Jupiter and the Earth*

The detection of radiation from Jupiter leads to the possibility of finding information concerning propagation conditions in interplanetary space near the plane of the ecliptic and especially along ray paths which pass near the Sun. Up to the present time the only information has come from the occultation of the discrete source Taurus-A by the Sun, a phenomenon studied in some detail by Hewish (1955). The large, and at present unpredictable, variations in intensity of the emission from Jupiter would seem to suggest that as a source of radiation for studying this problem Jupiter would not be satisfactory, but its appearance in bursts may actually be useful. In addition, its intensity is very high, frequently greater than that of any known discrete source, and a very small upper limit of about 1 min of arc is set to the angular size of the source. The observations described in this paper did not cover a very long period of time but in the course of the year there were noteworthy changes in the characteristics of the radiation from Jupiter.

Although no detailed study of the average daily intensity was made, it was apparent that during the period February-April 1951, when Jupiter's transit occurred near midday, the intensity was generally lower, by a factor greater than 3, than in the other periods, when transit occurred at night. Variations in ionospheric absorption would change the received intensity by a factor of less than 1·2 (see, for example, Mitra and Shain 1953) and the change in the Earth-Jupiter distance would account for a further factor of about 1·8. The remaining factor of 1·4 or more may have been due to the close approach of the ray paths to the Sun. If a simple, spherically-symmetrical, model of the corona is assumed, then at 18·3 Mc/s a source should be occulted by the solar corona at an angular distance of about $2^\circ{\cdot}2$ from the centre of the Sun (Bracewell and Preston, unpublished data). Jupiter was certainly detected (records being available for direction-finding) on March 7 at 13^h and on March 17 at 10^h, the distances of Jupiter from the centre of the Sun being then $3^\circ\cdot0$ and $4^\circ\cdot6$ (11 and 17 solar radii) respectively. In each case there was no evidence of refraction in the corona, within the experimental uncertainty. The record on March 9 showed features similar to those on March 7 and 17 and, although the direction-finding facility could not be used on this occasion, it was almost certainly radiation from Jupiter at a distance of $1^\circ\cdot6$ (6 solar radii) from the centre of the Sun. These results agree with those of Hewish, who also found that radiation could be observed at a time when, according to the simple model of the corona, the source should have been occulted. The fact that radiation could not be observed, although conditions were favourable, for 5 days after conjunction on March 12 suggests that there may have been asymmetry in the southern half of the outer corona.

A point of some interest is the changing appearance of the records, as shown in Figure 1. During October-November 1950 (and later in August-September 1951) the bursts were very short, with durations less than a minute. In February and April 1951 durations of several minutes were common, whilst for a few weeks near the time of conjunction the records often showed only slow variations such as can be seen on the record for March 17. Jupiter has a small angular size and might be expected to show rapid scintillations of large amplitude, arising in the passage of the radiation through the Earth's ionosphere, but none of the discrete cosmic sources has shown such marked variations in the characteristics of scintillations near midday.

The explanation of this lengthening of the bursts probably lies in the scattering of the radiation in the outer corona. Even on a simple model of the corona each radiated burst could reach the receiver by two paths, one direct and the other reflected from the Sun. As an example of the time differences involved, the separation in time of the direct and reflected rays would be about 2^s for an angular separation between the source and the centre of the Sun of 5°, with the delay nearly proportional to the square of the separation. As the refractive index of the corona is less than unity, there will be a further delay due to the slowing down of the reflected ray but this would be extremely small (of the order of $10^{-5\ s}$ at a separation of 5° and varying inversely as the fifth power of the separation). These delays would be too small to extend the bursts sufficiently, but much longer delays could arise due to such scattering processes as those considered by Hewish. A study of the short-period variations in the radiation from Jupiter should be a useful supplement to the work on the intensity variations of Taurus-A near occultation.

IV. Conclusions

As a result of the investigation described in this paper two facts have been clearly established. Firstly, the discovery by Burke and Franklin of radio radiation from Jupiter has been confirmed, and secondly, in 1951 the radio emission came from a very small portion of the planet. The active region is very probably identified with a visually disturbed region in the South Temperate Belt.

The study of Jupiter radiation should certainly be extended to observations at a number of frequencies. Such work would help, for example, in sorting out the effects of a possible Jovian ionosphere and might throw more light on the mechanism of origin of the radiation. It is also clear that, as a powerful source of short bursts of radiation, and with its known position and small angular size, Jupiter should be a useful object of study for obtaining information concerning the outer regions of the solar corona.

V. Acknowledgments

The author is indebted to Dr. B. F. Burke for the early communication of his results, which stimulated this investigation. Thanks are due to Dr. G. de Vaucouleurs of the Yale-Columbia Southern Station, and to Mr. J. Robertson of the Sydney Observatory, who gave much help in obtaining information concerning visual observations of Jupiter, and also to Dr. A. F. O'D. Alexander, the Director of the Jupiter Section of the British Astronomical Association, who in personal communications added further comments to the published work of the Section.

VI. References

Burke, B. F., and Franklin, K. L. (1955).—*J. Geophys. Res.* **60** : 213–7.
Fox, W. E. (1952).—*J. Brit. Astr. Ass.* **62** : 280–2.
Hewish, A. (1955).—*Proc. Roy. Soc.* A **228** : 238–51.
Higgs, A. J. (1951).—*Science News* **21** : 39–60.
du Martheray, M., and Antonini, E. (1952).—*Docum. Observateurs* **5**, No. 7.
Mitra, A. P., and Shain, C. A. (1953).—*J. Atmos. Terr. Phys.* **4** : 204–17.
Shain, C. A., and Higgins, C. S. (1953).—*Aust. J. Phys.* **7** : 130–49.

COSMIC RADIO SOURCES OBSERVED AT 600 Mc/s

By J. H. Piddington* and G. H. Trent*

[*Manuscript received October* 31, 1955]

Summary

A survey of cosmic radio emission at 600 Mc/s using a beam 3°·3 wide has been made over most of the celestial sphere between declinations 90 °S. and 51 °N. Discrete sources numbering 49 were located having flux densities of about 1×10^{-24} W m^{-2} $(c/s)^{-1}$ or more. Of these, 18 do not appear to have been reported previously, including 12 which lie within ±2° of the galactic plane.

Most of the sources seem to be associated with other sources or with irregularities of the background radiation. There seems to be no sharp boundary between discrete sources and such irregularities. Many of the new sources are likely to be clouds of thermally emitting H II.

I. Introduction

A survey of cosmic radio emission at 600 Mc/s has been made over most of the celestial sphere between declinations 90 °S. and 51 °N. from a field station near Sydney, N.S.W. This communication presents the data on the localized sources found during the survey.

The results have been obtained using a narrow pencil beam of width 3°·3 between half-power points. Although narrower beams have been employed, the observations have generally been confined to limited regions of the sky and even the most extensive pencil beam surveys have not included the southern Milky Way, a region rich in radio sources. The most comprehensive surveys of discrete sources have been made with interferometer techniques at frequencies of the order of 100 Mc/s. The use of a beam at a much higher frequency tends to highlight spread sources and sources with unusual spectral characteristics such as 20N4A (Cygnus-X, Piddington and Minnett 1952). A number of new sources of large angular width and with peculiar spectra have been found. The spectral characteristics will be discussed in another communication.

II. Experimental Methods

(*a*) *The Aerial System*

The aerial system used was a paraboloid 36 ft in diameter used as a transit telescope. It has previously been described (Kerr, Hindman, and Robinson 1954) in connexion with hydrogen line observations at 1420 Mc/s. The feed system was a single dipole and reflector plate. A direct measurement of aerial gain has not been made, but it is unlikely that such a measurement could improve greatly on the estimate made as follows : using the Sun as a steady, and quite

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

powerful, source of radiation a plot was made of the power response in the main aerial beam. The finite angular size of the Sun is not sufficient to change the plot appreciably from that given by a point source. If we neglect side and back aerial lobes for the time being, the "nominal gain" over an isotropic radiator may be determined with an accuracy of a few per cent. by a process of graphical integration.* This was found to be 3800. The real gain was then found from an estimate of the proportion of radiated power in the main beam (the aerial assumed transmitting), which was 0·65. The overall aerial gain was therefore 2500.

The relationship between the aerial declination setting and true declination was found to vary with the setting, owing to changes in the sag of the structure. A correction table derived by Kerr, Hindman, and Robinson was used; this was based mainly on visual sights of stars, using radio observations on the Sun to relate the mechanical and radio axes of the aerial.

The directional accuracy for strong, isolated sources is about ½° in both Right Ascension and declination. However, when the source is weak or is close to other emitting regions, the accuracy is reduced.

A second, smaller aerial (6 ft diameter) was used in conjunction with the large aerial to act as a temperature reference level. The receiver then measured the difference in flux received by the two aerials. The second aerial was pointed at the south celestial pole so as to receive a steady signal in its main beam. The passage of parts of the sky with high brightness temperatures through the side lobes has a negligible effect so that the total received signal does not vary appreciably. Its feeder system was adjusted to have the same loss (3 dB) and to be subjected to approximately the same temperature fluctuations (due to ambient temperature changes) as the main feeder.

(*b*) *The Receiver*

The receiver operated on the Dicke system of switching at 25 c/s between the two aerial feeders. Apart from changes necessitated by the new operating frequency most of the equipment is similar to that already described (Piddington and Minnett 1951) for operation at 1210 Mc/s. A determination of the receiver sensitivity was made before and after each series of observations, using a matched termination in place of the aerial. The termination could be heated through a known temperature range.

(*c*) *Observations*

Observations were made by setting the main aerial to a given declination and recording the level of intensity over a period of several hours. For the most part satisfactory conditions occurred only between about 2200 hours and 0600 hours local time. Occasionally the main aerial was pointed at the South

---

* It should be noted that this is purely a graphical process involving no data other than the relative response of the aerial in different directions. The physical basis may be better seen by imagining the aerial transmitting the same amount of power as an isotropic radiator. If the aerial response is such that the power radiated in the centre of the main beam is n times the average over the whole sphere, then the aerial gain is n.

Pole to give a fixed reference level. This level could not be assumed a zero level for other declinations, however, because aerial side lobes picked up different amounts of ground radiation for different aerial positions. The zero level for each declination was taken as the " cold " part of the sky, well clear of the Milky Way. Extrapolating the measured spectrum of the cold sky gives a brightness temperature of about 4 °K at 600 Mc/s and suggests that the method used is quite safe provided that it is remembered that all brightness temperatures measured are above an unknown minimum of a few degrees.

The receiver output was automatically recorded, together with Right Ascension noted at hourly intervals. An example of a record made at Dec. 49°·9 S., R.A. between about $13^h\ 00^m$ and $19^h\ 00^m$ is shown in Figure 1. This shows the powerful source in Centaurus and later the galactic crossing which provide receiver input temperatures of about 10 and 50 °K respectively. This record is discussed below.

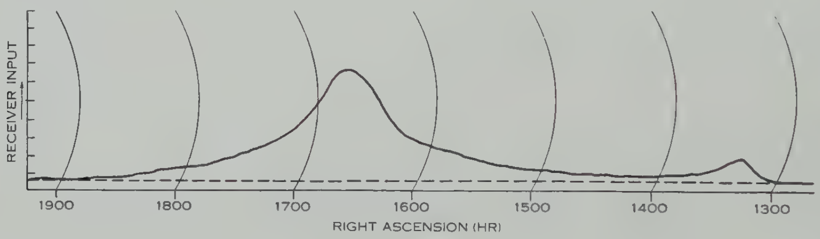


Fig. 1.—Record taken at Dec. 44°·9 S. between R.A. $13^h\ 00^m$ and $19^h\ 00^m$. The Centaurus source (13S4A) and its " link " with the main galactic radiation is shown. The receiver input temperature is about 8 °K per division.

The receiver sensitivity was measured before and after each record by means of a matched load which could be heated to temperatures of the order 100 °C. It was found that under ideal conditions sources giving flux densities of 10^{-24} W m^{-2} $(c/s)^{-1}$ could be detected with reasonable ease.* However, many records were far from perfect, owing to interference, and even after several repetitions good records were not obtainable for some parts of the sky. It may be assumed that most of the region within about 30° of the Milky Way was satisfactorily surveyed except beyond declinations 51 °N. and 70 °S. The region south of 70° is difficult to survey because of the slow rate of transit of any sources ; this accentuates the effects of random drifts. However, some runs were made through the Magellanic Clouds and the results are given.

III. Results

In Table 1 are listed 49 sources observed at 600 Mc/s. Previous lists of radio sources have frequently described them as " discrete ", implying the existence of an isolated physical entity. It is unlikely that all of the present sources are discrete in that sense ; many of them, particularly those lying near the galactic plane, appear on the records as localized maxima above strong background radiation. Their degree of association with the background radiation cannot be determined but each represents a reasonably well-defined maximum of intensity.

---

* This corresponds to a beam brightness temperature of 2·8 °K.

Also listed is the flux density of solar radiation on three occasions. This is included in the hope that the quiet Sun* may soon be used as a standard source of radiation by means of which different workers may check their aerial and receiver characteristics. The same consideration applies to the Moon (at higher frequencies) and to outstanding cosmic radio sources whose spectra may soon be known.

The sources are given a reference number and, when available, an International Astronomical Union number (Pawsey 1955). When they appear to coincide with a previously observed source for which no I.A.U. number is available the observer's initials and own reference number are given. These are as follows: HBH—Hanbury Brown and Hazard (1953); M—Mills (1952); BWSS—Bolton *et al.* (1954); AG—Allen and Gum (1950); BSS—Bolton, Stanley, and Slee (1954); KKM—Kraus, Ko, and Matt (1954); DLS—Denisse, Leroux, and Steinberg (1955); NRL—Haddock, Mayer, and Sloanaker (1954); M(1955)—Mills (1955).

For spread sources the constellation listed is that of the highest intensity level, when available, or of the centre. The approximate dimensions (R.A. × Dec.) are given to the contours of lowest intensity of these sources. When only one dimension is given this is the largest, the other being not greater than the aerial beamwidth. For spread sources the total flux density is found only by a process of integration. For this reason the value given is the maximum aerial temperature: this value is obtained after subtracting the mean background aerial temperature.

IV. Notes on the Sources

Source 1.—Mills (1954) working at 85·5 Mc/s has identified a source at R.A. $03^h\,20^m{\cdot}6$, Dec. $-37°\,23'$ with NGC 1316. This is probably the same source, although the difference in position, which is near the limit of experimental error, suggests that there may be a contribution from the surrounding region which has a different spectrum.

Source 2.—Near limit of sensitivity; seems spread ~5° in direction of changing declination. Agrees in position with HBH.8.

Source 3.—Near limit of detection; seems spread in direction of changing R.A.

Source 7.—This is probably the giant emission nebula 30 Doradus at R.A. $05^h\,40^m$, Dec. $-69°{\cdot}1$, lying within the Large Magellanic Cloud. This identification was made by Mills (1955) using a narrower (~1°) beam; he observed a strong maximum at R.A. $05^h\,40^m$, Dec. $-69°{\cdot}3$. The different radio positions

---

* As we move towards the next sunspot maximum the " quiet Sun flux density " (that is, the minimum flux density over a period of a few days or weeks) will increase steadily. Intercomparison of measurements made at different times can be made by using the daily values of solar flux density published in the *Quarterly Bulletin on Solar Activity* published by the International Astronomical Union.

TABLE 1

LIST OF SOURCES

Reference No.	Designation or Previous Identification	Constellation	Position of Centre				Approx. Angular Size (°)	Max. Aerial Beam Temp. (°K)	Flux Density (both polarizations × 10^{-24} W m^{-2} (c/s)$^{-1}$)	Remarks
			Celestial Coordinates (Epoch 1950)		Galactic Coordinates					
			R.A. (h m)	Dec. (°)	l (°)	b (°)				
1	03S3A	Fornax	03 23	−37½	206	−55	—	—	3·1	Probably NGC 1316 (Mills 1954)
2	HBH.8	Perseus	04 00	+48	120	− 2	∼5	∼3	—	Weak, spread source
3	——	Perseus	04 02	+40	126	− 8	∼5	∼3	—	Very weak and appears spread
4	M04+1	Orion	04 50	+10	157	−18	—	—	1·0	
5	05N2A	Taurus	05 30	+22	152	− 4	—	—	12·5	NGC 1952 (Crab Nebula)
6	05S0A	Orion	05 32	− 5½	177	−18	—	—	3·0	NGC 1976 (Orion Nebula)
7	M(1955)	Dorado	05 37	−69	246	−31	—	—	3·4	Probably 30 Doradus (Mills 1955)
8	——	Taurus	05 40	+28	151	0	—	—	—	Not seen on later runs
9	BWSS.D	Orion	05 42	− 1	174	−14	∼5×5	7	—	Spread source
10	——	Orion	05 50	+22	154	− 1	∼12	7	—	Spread source
11	06N2A	Gemini	06 13	+23	156	+ 4	—	—	2·2	IC 443
12	——	Monoceros	06 30	+ 6	173	0	10×20	7	—	Suggests part of galactic structure
13	——	Monoceros	06 33	+ 5½	174	0	—	—	4·0	May be NGC 2237
14	——	Gemini	07 15	+27	159	+19	—	—	∼1	Poor record
15	08S4A	Puppis	08 20	−42	227	− 2	—	∼5	—	
16	AG	Vela	08 36	−45½	232	− 2	18×15	56	—	Large galactic feature
17	BSS.119	Vela	09 10	−41½	233	+ 5	—	∼8	—	
18	BWSS.G	Hydra	09 15	−10	209	+27	—	—	1·5	
19	09S1A	Hydra	09 18	−11½	211	+27	—	—	0·8	
20	——	Leo	09 45	+19	182	+49	5×0	5	—	Poor record
21	——	Carina	10 49	−59½	256	− 1	6×0	39	—	
22	——	Centaurus	11 40	−58½	262	+ 2	—	∼3	—	

23	——	Crux	12 10	−62	266	0	—	~3	—	
24	12N1A	Virgo	12 29	+12½	261	+74	—	—	3·6	NGC 4486
25	——	Centaurus	12 38	−64½	269	− 3	—	~6	—	
26	——	Centaurus	13 18	−63	274	− 1	—	~11	—	
27	13S4A	Centaurus	13 22	−43	278	+18	—	—	13	NGC 5128 plus spread source
28	13S6A	Centaurus	13 40	−60	277	0	—	~6	—	
29	——	Centaurus	14 07	−61½	280	− 1	—	~30	—	
30	KKM.Hya0	Hydra	14 50	−29	301	+25	—	—	1	
31	——	Circinus	14 58	−59	287	− 2	—	14	—	
32	16S6A	Norma	16 08	−60	293	− 7	—	5	—	
33	——	Ara	16 40	−46	307	− 2	—	22	—	May contain NRL 19
34	16N0A	Hercules	16 45	+ 6	351	+29	—	—	1	
35	M17–3	Scorpius	17 10	−39	315	− 1	—	10	—	
36	NRL12	Scorpius	17 25	−35	320	− 2	—	25	—	
37	17S2A	Sagittarius	17 44	−29	328	− 1	—	—	29	Has been proposed as galactic nucleus
38	DLS	Ophiuchus	17 45	+ 7½	360	+16	6	11	—	Spread source
39	——	Scorpius	17 50	−45	314	−11	6	5	—	Spread source
40	NRL10 & 11	Sagittarius	18 01	−22½	335	− 2	—	—	7	
41	NRL4	Sagittarius	18 14	−16½	342	− 2	—	—	4	Perhaps Omega Neb. NGC 6618
42	——	Hercules	18 26	+18	14	+12	—	5	—	
43	KKM.Sct.A	Scutum	18 32	− 8	352	− 1	—	—	4	
44	M19+0	Aquila	19 04	+ 8	9	− 1	—	—	6	
45	——	Aquila	19 22	+13½	16	− 2	—	—	3	
46	19N4A	Cygnus	19 58	+40½	44	+ 5	—	—	37	Original Cygnus–A source
47	20N4A	Cygnus	20 24	+40½	46	+ 1	6	90	—	Original Cygnus–X source
48	HBH.21	Cygnus	20 49	+51	57	+ 4	—	—	~3	
49	NRL21	Cygnus	20 57	+44¾	54	− 1	—	~5	—	Suggested identification NGC 7000
Sun	——	——	—	—	—	—	—	—	3100	Feb. 1, 1955
Sun	——	——	—	—	—	—	—	—	3900	Sept. 13 and 14, 1955

are no doubt caused by the spread of bright areas around 30 Doradus ; the flux density received from these areas depends on the aerial beamwidth and the spectral characteristics. The emission is probably thermal, from ionized hydrogen, and the maximum lies within 1° of that for line emission by neutral hydrogen observed by Kerr, Hindman, and Robinson (1954).

Source 9.—A clearly defined spread source whose contours may just adjoin those of the Orion Nebula (source 6).

Source 10.—This appears to be a source spread $\frac{1}{2}$–1 hour in Right Ascension and perhaps a couple of degrees in declination. It may merely be the two sources recognized within the area (Crab Nebula and IC 443) with a little additional radiation from another source lying between them.

Source 12.—This large, weak source lying along the galactic plane between $l \sim 165°$ and $l \sim 185°$ may be a substantial part of the galaxy. It does not show very close agreement with the contours of Reber (1948) for 480 Mc/s radiation ; these extend from $l \sim 143°$ to $l \sim 180°$.

Source 13.—This is a source within the spread source 12. It may be identified with NGC 2237 (Mills, personal communication).

Source 16.—With the sensitivity and aerial beam of the 600 Mc/s radiometer this emission appears as a large but completely isolated source, roughly circular in shape. It extends along the galactic plane from $l \sim 222°$ to $l \sim 244°$ with maximum intensity at $l = 232°$, $b = -2°$. Its existence and approximate location were inferred by Allen and Gum (1950) from measurements at 200 Mc/s with a 25° wide aerial beam ; they did not separate it from the general galactic emission. It is also a very conspicuous object at 85·5 Mc/s (Mills, personal communication). Finally, it approximately coincides in position with one of the main regions of 1420 Mc/s H I line emission (Christiansen and Hindman 1952).

Source 21.—This is a source spread about 6° along the galactic plane but having no measurable width across the plane. Although a very conspicuous object at 600 Mc/s it has not been reported previously ; high frequency surveys have not included this portion of the sky and the object is not outstanding at lower frequencies. At 85·5 Mc/s, Mills (personal communication) finds several rather weak sources of small angular size which would probably not be resolved by the 3° beam.

Sources 22, 23, 25, 26, 28, 29, 31, 32, 36, 37, 40, 41, 43, 44, 45.—These sources may be discrete or may be maxima of the general galactic radiation. They are all in regions where the background is bright and non-uniform. Their apparent size is consistent with point sources but they might have dimensions up to a few degrees.

Source 32.—The position found is in Norma but the position of 16S6A is just across the border in Triangulum Australis.

Source 35.—The previous identification given M17–3 may not refer to the same source, or combination of sources, since it was at R.A. $17^{\mathrm{h}}\ 20^{\mathrm{m}}$, Dec. −39° (about 2° away). McGee, Slee, and Stanley (1955) found a source at R.A. $17^{\mathrm{h}}\ 13^{\mathrm{m}}$, Dec. −38° which is closer, and Haddock, Mayer, and Sloanaker (1954) found one at R.A. $17^{\rm h}\ 17^{\rm m}$, Dec. −36° which is further away. There may be a combination of sources in the region.

Source 38.—This spread source may be discrete or part of the background to which it appears to be attached. It is evident in the contour plots of Denisse, Leroux, and Steinberg (1955) and Kraus and Ko (1955).

Source 40.—This is probably M20 which is at R.A. $17^{\rm h}\ 59^{\rm m}$, Dec. −23°, together with radiation from nearby objects including M8. The identification has been made at 3200 Mc/s by Haddock, Mayer, and Sloanaker (1954).

Source 46.—There may be several sources here. Hanbury Brown and Hazard (1953) report a source at R.A. $20^{\rm h}\ 44^{\rm m}$ and Denisse, Leroux, and Steinberg (1955) a source at R.A. $20^{\rm h}\ 57^{\rm m}$; all declinations are about $50^\circ\frac{1}{2}$.

Sun on February 1, 1955.—The flux density found here should not be compared with the 600 Mc/s Sydney measurements given in the *Quarterly Bulletin on Solar Activity*, since the latter are unreliable as far as absolute level is concerned (Smerd, personal communication). If we take into account the increase in flux with frequency it accords well with the 545 Mc/s Nederhorst result for the same day of 2900 units of flux compared with our 3100 units.

V. Variation of Intensity Along the Galactic Plane

Most of the newly discovered sources lie on or near the galactic plane. It is possible, therefore, to convey an impression of their relative positions, extent, and emission strength by plotting receiver input against galactic longitude for zero galactic latitude. This is done in Figure 2 for longitudes between 210 and 360°, the various sources being marked by their reference numbers listed in Table 1. The maxima which do not lie exactly on the galactic plane but within a few degrees of this plane are also shown, by dotted lines. These

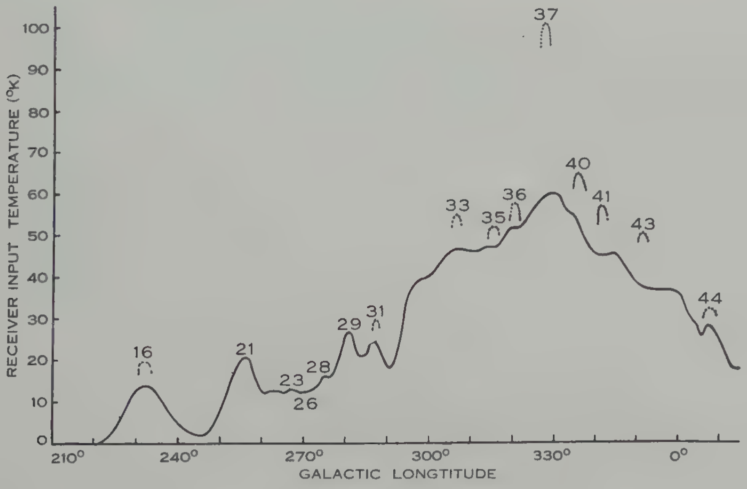


Fig. 2.—Variation of intensity along the galactic plane. Maxima (sources) within a few degrees of the plane (but not on it) are shown by dashed lines.

include the outstanding maximum at $l=328^\circ$ in the direction of the galactic centre. When a source lies near the galactic plane and partly overlaps it (for example, source 44) it appears as a dotted maximum above a maximum in the full curve.

Figure 2 shows the concentration of sources near the galactic plane from which it may be inferred that they have a galactic origin, possibly in H II regions. The sources are nearly as numerous in the southern portion of the galaxy as towards the galactic centre. This provides general evidence of an origin in H II regions, as these are known to be numerous in the southern sky.

VI. Discussion

Early observations of radio sources were made using interferometers having numerous very narrow beams. That they were observed at all indicated that the sources were of limited angular extent and it was generally assumed that they were discrete. Subsequently some sources were found to be associated with irregularities in the background emission. A noteworthy feature of the present survey is that nearly all the sources are associated with such irregularities, often to a high degree. Thus in many cases a more apt term than " discrete sources " might be localized maxima. The effect is more apparent at high than at low frequencies, although, of course, any association between different radiating regions is independent of frequency. The apparent enhancement at high frequencies must result from different emission (and perhaps absorption) spectra of different regions.

Table 2

POSSIBLE ASSOCIATIONS OF RADIO AND VISIBLE SOURCES

Radio Source	Nebula BBW	Celestial Coordinates 1955		Dimensions (m)
		R.A. (h m)	Dec. (°)	
21	25500	10 43	$-59\frac{1}{2}$	450×330
26	27500	13 30	-62	45×45
29	28100	14 07	-61	61×83
31	28802	15 12	$-59\frac{1}{4}$	39×51

An example of association is source 27 (13S4A) in Centaurus which has been identified with the extragalactic object NGC 5128. An extended source, about 2° in diameter and concentric with the original source, has been found by Bolton *et al.* (1954). Our source must include both of these and also shows background emission from a narrow region which joins the main galactic emission some 15° away from the source. The linkage with the Galaxy is evident in the record reproduced in Figure 1.

The survey shows 18 sources which do not appear to have been listed previously. Of these, 12 lie within ±2° of the galactic plane. It is likely that most of the latter are thermally emitting clouds of ionized hydrogen. This is suggested by their proximity to the galactic plane and by the fact that they show up better at the relatively high frequency used. There is some evidence that four of the new sources are at least associated with visually observed H II regions.

VII. Possible Associations

A superficial comparison of the radio results with visual observations of southern H II regions made by Bok, Bester, and Wade (1954) has been made and four likely associations found. These are listed in Table 2.

A much more detailed comparison is necessary before any actual identification of radio and visual emitting regions is made.

VIII. References

Allen, C. W., and Gum, C. S. (1950).—*Aust. J. Sci. Res.* A **3** : 224.

Bok, B. J., Bester, M. J., and Wade, C. M. (1954).—Catalogue of H II regions in the Milky Way. Harvard Observatory and Boyden Station Rep.

Bolton, J. G., Stanley, G. J., and Slee, O. B. (1954).—*Aust. J. Phys.* **7** : 110.

Bolton, J. G., Westfold, K. C., Stanley, G. J., and Slee, O. B. (1954).—*Aust. J. Phys.* **7** : 96.

Brown, R. Hanbury, and Hazard, C. (1953).—*Mon. Not. R. Astr. Soc.* **113** : 123.

Christiansen, W. N., and Hindman, J. V. (1952).—*Aust. J. Sci. Res.* A **5** : 437.

Denisse, J. F., Leroux, E., and Steinberg, J. L. (1955).—*C.R. Acad. Sci., Paris* **240** : 278.

Haddock, F. T., Mayer, C. H., and Sloanaker, R. M. (1954).—*Nature* **174** : 176.

Kerr, F. J., Hindman, J. V., and Robinson, B. J. (1954).—*Aust. J. Phys.* **7** : 297.

Kraus, J. D., and Ko, J. C. (1955).—*Nature* **175** : 159.

Kraus, J. D., Ko, J. C., and Matt, S. (1954).—*Astr. J.* **59** : 439.

McGee, R. X., Slee, O. B., and Stanley, G. J. (1955).—*Aust. J. Phys.* **8** : 347.

Mills, B. Y. (1952).—*Aust. J. Sci. Res.* A **5** : 266.

Mills, B. Y. (1954).—*Observatory* **74** : 248.

Mills, B. Y. (1955).—*Aust. J. Phys.* **8** : 368.

Pawsey, J. L. (1955).—*Astrophys. J.* **121** : 1.

Piddington, J. H., and Minnett, H. C. (1951).—*Aust. J. Sci. Res.* A **4** : 459.

Piddington, J. H., and Minnett, H. C. (1952).—*Aust. J. Sci. Res.* A **5** : 17.

Reber, G. (1948).—*Proc. Inst. Radio Engrs., N.Y.* **36** : 1215.

RADIO EMISSION FROM NOVAE AND SUPERNOVAE

By B. Y. Mills,* A. G. Little,* and K. V. Sheridan*

[*Manuscript received November* 7, 1955]

Summary

Attempts have been made to observe the radio emission at 3·5 m from two supernovae and ten novae. Kepler's star was the only reasonably certain identification. A comparison with radio observations of other supernova remnants suggests a constant ratio between the present radio emission and the maximum emission of light. It is concluded that for common novae, which are not detectable as radio sources, this ratio must be smaller than for supernovae. The galactic radio emission near the plane of the Milky Way could be largely the integrated emission of supernova remnants but common novae could not contribute appreciably.

I. Introduction

In the course of the initial tests on the large 3·5 m radio telescope now in operation near Sydney, attempts were made to observe a considerable number of known celestial objects. Of these observations some, which have already been described (Mills 1955), led to general conclusions regarding the distribution of the galactic radio emission. Two components were identified, one having a discoidal distribution strongly concentrated towards the galactic plane and the other being nearly spherical and forming a " corona " surrounding the Galaxy. A series of observations were made in an attempt to elucidate the nature of the discoidal distribution by means of a direct study of two classes of object which there is reason to suppose might be associated with the radio emission, that is, novae, including supernovae, and emission nebulae. It is found however, that the data are not yet sufficient for a definite answer to this question, which will require a detailed study of the distribution of radio emission over large areas near the galactic plane preferably at more than one frequency. The results of a current survey of the southern sky should provide much of the necessary data ; meanwhile the present observations, which are described in this and a companion paper (Mills, Little, and Sheridan 1956), lead to some conclusions of interest.

It is well known that the intense radio source 05N2A may be identified with the Crab Nebula, which is the visible remnant of the supernova of 1054 A.D. This identification was first suggested by Bolton, Stanley, and Slee (1949) and has now been well established as the result of several independent investigations. A radio source has also been found close to the position of another well-known nova, that of Tycho Brahe (Hanbury Brown and Hazard 1952). Although no visible remnants of this nova have yet been discovered and, accordingly, a

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

positive identification cannot be made, the positions are in agreement and, moreover, there is good reason for believing that this also was a supernova (Baade 1938). In addition, at the position of some other sources, notably 23N5A and 06N2A, nebulosities exist which could well be the remnants of supernovae.* More recently, also, a radio source of large angular size concentric with the Loop Nebula in Cygnus was detected by Walsh and Hanbury Brown (1955); it has been suggested by Oort (1946) that this nebula is a supernova remnant.

These results suggest very strongly that supernova outbursts lead to prolonged and intense radio emission from the gaseous remnants. It has even been suggested by Hanbury Brown (1954) and Shklovskii (1954) that all the non-thermal class I radio sources are supernova remnants. Shklovskii has attempted to identify the stronger sources with various ancient novae which he regards as supernovae, while Hanbury Brown has shown that, by making plausible estimates of the frequency of occurrence of supernovae in the Galaxy and the duration of their radio emission, it may be possible to account for the observed metre-wavelength radio emission near the galactic plane by their integrated emission. An extension of these ideas to ordinary novae is suggested by the identification of Nova Aquila 1918 with a strong radio source by Bolton, Stanley, and Slee (1954) but, as we shall see later, this identification is incorrect.

II. OBSERVATIONS

As a result of these considerations a number of special observations were made at the positions of bright novae and supernovae. In addition, records made for other purposes were searched for evidence of radio emission from fainter novae and from ancient novae of doubtful positional accuracy. The results are shown in Table 1, with the exception of those of the ancient novae which will be discussed later; data derived from published observations of northern novae are also included. The magnitude scale of Hanbury Brown and Hazard (1952) is used for specifying the radio measures, i.e. $m_{R(3\cdot 5\mathrm{m})} = -53\cdot 4 - 2\cdot 5 \log_{10} S_{(3\cdot 5\mathrm{m})}$. The optical magnitudes m have been taken from standard sources (e.g. Norton and Inglis 1943; Pawsey and Bracewell 1955); some are photographic and some visual magnitudes but the accuracy required does not warrant differentiating between them. No attempt is made to correct the radio magnitudes to a standard wavelength as the spectrum of these sources is not sufficiently well established. Accordingly the radio magnitude of Tycho Brahe's nova refers to the wavelength at which the measurement was made, that is, 1·9 m.

Since all the present observations have been made close to the galactic plane the sensitivity of the radio telescope is much less than before (Mills 1955) when the external galaxies under investigation were at relatively high galactic latitudes. The reduction in sensitivity is caused, firstly, by the higher brightness temperatures of the background radiation, which increase the "noise level" of the equipment and, secondly, because of the greater irregularities in the distribution and the higher source density, for which the angular resolution of the equipment is often inadequate.

---

* Originally it was thought that the nebulosity associated with 23N5A could not be a supernova remnant (Baade and Minkowski 1954), but as a result of further observations this possibility is now admitted by Minkowski (reported at the Jodrell Bank symposium on radio astronomy, 1955).

The identification of Kepler's nova with a radio source may be considered as reasonably certain. Tracings of records obtained on three declinations straddling the nova are shown in Figure 1. The source position is $17^{\mathrm{h}}\,26^{\mathrm{m}}{\cdot}8$, $-21^\circ\,29'$ (1935) with a probable error of about $0^{\mathrm{m}}{\cdot}1$ in Right Ascension and $3'$ in declination; compared with the nova position of $17^{\mathrm{h}}\,26^{\mathrm{m}}\,44^{\mathrm{s}}{\cdot}9$, $-21^\circ\,25'\,55''$ (1935) and a position of $17^{\mathrm{h}}\,26^{\mathrm{m}}\,42^{\mathrm{s}}{\cdot}8$, $-21^\circ\,25'\,54''$ (1935) for a fragment of the gaseous remnant observed by Baade (1943). The identifications of three of the supernovae with radio sources are further supported by the remarkable agreement in the values of $m_R - m$, indicating an almost constant ratio between their present radio emission and their maximum light emission. This agreement is not as good as it appears, however, because of the probability that the optical obscuration is different in each case.

TABLE 1

NOVA OBSERVATIONS

Nova	Flux Density at 3·5 m (W $m^{-2}(c/s)^{-1}$ $\times 10^{-26}$)	$m_{R(3\cdot5\mathrm{m})}$	m at Maximum	$m_{R(3\cdot5\mathrm{m})}-m$	Remarks
Southern Novae					
1604 Ophiuchi	38	7·5	−2	$9\frac{1}{2}$	Kepler's star, a supernova
1860 Scorpii ..	<10	>9	7	>2	
1895 Carinae	<10	>9	8	>1	
1895 Centauri	<5	>10	7	>3	A supernova in NGC 5253
1898 Sagittarii	<5	>10	4·7	$>5\frac{1}{2}$	
1899 Sagittarii	<10	>9	8·5	$>\frac{1}{2}$	
1899 Aquilae ..	<10	>9	7	>2	
1910 Sagittarii	<20	$>8\cdot3$	7·5	$>\frac{1}{2}$	
1917 Ophiuchi	<10	>9	6·5	$>2\frac{1}{2}$	
1918 Aquilae ..	<10	>9	−0·7	$>9\frac{1}{2}$	Wrongly identified as a strong radio source by Bolton, Stanley, and Slee (1954)
1925 Pictoris ..	<5	$>9\cdot7$	1·1	$>8\frac{1}{2}$	
1942 Puppis ..	<5	$>9\cdot7$	0·4	$>9\frac{1}{2}$	
Northern Novae					
1054 Tauri ..	1800	3·4	−6	$9\frac{1}{2}$	The Crab Nebula, a supernova
1572 Cassiopeiae	170*	6·0*	−4	10*	Tycho Brahe's star, a supernova

* Measured at a wavelength of 1·9 m.

Most of the common novae were too faint for their radio emission to be detected if the same proportionality between radio and optical emission were preserved. However, the three brightest do give significant minimum values of $m_R - m$, and for these it may be concluded that the ratio of present radio emission to brightness at maximum does not exceed the corresponding ratio for supernovae. There is no detectable excess emission at the position of Nova Aquila 1918; the source which Bolton, Stanley, and Slee identified with the nova is probably an intense extended source about 1° away in a north-following direction.

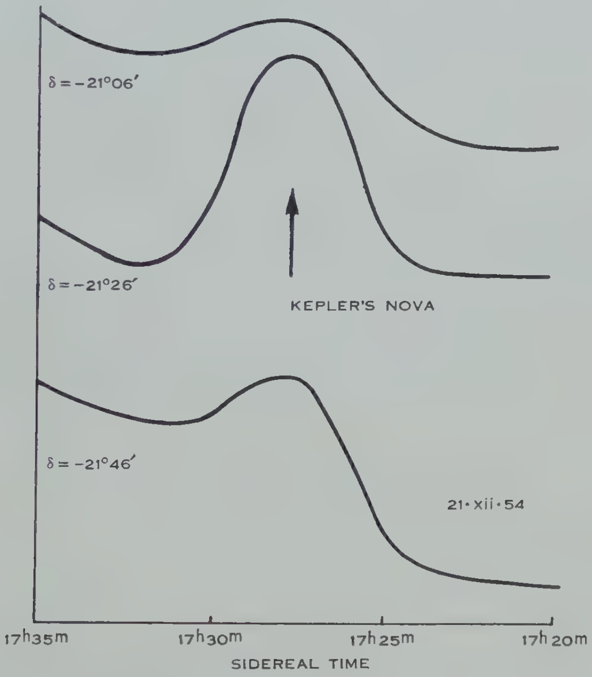


Fig. 1.—Tracings of three records at declinations near Kepler's nova.

Attempts made by Shklovskii (1954) to identify some strong radio sources with ancient novae are not convincing. Records taken in these regions with the present equipment generally reveal a number of sources within the uncertainty in position of the nova and there seems little justification for assuming that an identification can always be made with the strongest: progress must be dependent on the discovery of gaseous remnants at the position of the radio source. At least one of his suggested identifications is not supported by the accurate radio position now available, that is, the identification of the nova of 185 A.D. in Centaurus with the radio source 13S6A. Lundmark (1921) places the nova between α and β Centauri, whereas the position of the radio source is $13^h\,43^m{\cdot}0$, $-60°\,12'$ (1950) with probable errors of $0^m{\cdot}2$ in Right Ascension and $3'$ in declination: this is well outside Lundmark's limits.

III. DISCUSSION

The most surprising result of these observations is the agreement between the values of $m_R - m$ for Kepler's star and the two northern supernovae. Thus the subsequent radio emission of a type I supernova outburst, of which these are examples, appears to share in the uniformity of the optical spectrum and decay curve ; it would be very interesting to compare the spectra of the three radio sources. Shklovskii (1953) has suggested that the optical emission of the Crab Nebula has the same origin as the radio emission and that both are caused by the emission from relativistic electrons moving in a magnetic field ; this suggestion has now received substantial support from the measurement of strong linear polarization in the optical emission of the nebula (reported by Oort at the Jodrell Bank symposium on radio astronomy, 1955). It appears that the so-called " amorphous mass " of the nebula, which has a continuous spectrum, actually consists of a cloud of relativistic electrons. Despite the similar radio emission of the other supernova remnants, there is no corresponding similarity in their optical emission. Neither Kepler's nova, which has a remnant displaying a typical filamentary structure with strong emission lines, nor Tycho Brahe's nova, which has no visible remnant, appears to radiate at optical wavelengths by this mechanism. It would seem that, either their associated magnetic fields are much smaller than in the case of the Crab Nebula or, more probably, electron energies do not reach such high values. If electrons are being accelerated by a Fermi process in these remnants they may yet become visible.

The contribution of supernova remnants to radio emission near the galactic plane has been considered in detail by Hanbury Brown (1954). Quite apart from the correctness of the mechanism he assumes, his arguments show that, if their period of activity is 10^5 years and their rate of occurrence in the Galaxy about once in 20 years, corresponding to 5×10^4 active remnants, it is possible to account for the discoidal distribution by their integrated emission. The occurrence rate assumed is some 15 times that derived by Zwicky (1942) from a study of external galaxies, but is not inconsistent, since Zwicky found that some galaxies may be very prone to supernova outbursts and the Milky Way may be such a galaxy.*

The failure to detect radio emission from any common nova in the present observations may be used to show that their integrated emission is unlikely to be of any significance. Assuming that $m_R - m$ for a nova is greater than that for a supernova, the radio emission from each must accordingly differ by a factor greater than the difference in their light emission at maximum brightness, that is, by a factor of about 10^4. The rate of occurrence of novae in the galaxy is probably of the order of 10 per year, so that, if the lifetime of their radio emission does not exceed that of supernovae, they can at most contribute a few per cent. of the total radio emission near the galactic plane.

---

* *Note added in Proof.*—In discussion with the authors, Minkowski has remarked that these differences are explicable if account is taken of the two forms of supernovae, type I associated with Population II and of very low occurrence frequency and type II associated with Population I and very frequent in Sb and Sc galaxies. Since the supernovae we discuss are all of type I, with an average recurrence period of hundreds of years, we may conclude that the integrated radio emission of all such supernova remnants in the Milky Way is likely to be insignificant. However, the possibility remains that type II supernova remnants, of which the intense radio source in Cassiopeia seems to be an example, may contribute significantly to the radio emission close to the galactic plane.

IV. REFERENCES

BAADE, W. (1938).—*Astrophys. J.* **88** : 285.
BAADE, W. (1943).—*Astrophys. J.* **97** : 119.
BAADE, W., and MINKOWSKI, R. (1954).—*Astrophys. J.* **119** : 215.
BOLTON, J. G., STANLEY, G. J., and SLEE, O. B. (1949).—*Nature* **164** : 101.
BOLTON, J. G., STANLEY, G. J., and SLEE, O. B. (1954).—*Aust. J. Phys.* **7** : 110.
BROWN, R. HANBURY (1954).—*Observatory* **74** : 183.
BROWN, R. HANBURY, and HAZARD, C. (1952).—*Phil. Mag.* **43** : 123.
LUNDMARK, K. (1921).—*Publ. Astr. Soc. Pacif.* **33** : 225.
MILLS, B. Y. (1955).—*Aust. J. Phys.* **8** : 368.
MILLS, B. Y., LITTLE, A. G., and SHERIDAN, K. V. (1956).—Emission nebulae as radio sources. *Aust. J. Phys.* **9** (2) (in press).
NORTON, A. P., and INGLIS, J. G. (1943).—" A Star Atlas and Reference Handbook." (Gall and Inglis : London.)
OORT, J. H. (1946).—*Mon. Not. R. Astr. Soc.* **106** : 159.
PAWSEY, J. L., and BRACEWELL, R. N. (1955).—" Radio Astronomy." (Clarendon Press : Oxford.)
SHKLOVSKII, I. S. (1953).—*C.R. Acad. Sci. U.R.S.S.* **90** : 983.
SHKLOVSKII, I. S. (1954).—*C.R. Acad. Sci. U.R.S.S.* **94** : 417.
WALSH, D., and BROWN, R. HANBURY (1955).—*Nature* **175** : 808.
ZWICKY, F. (1942).—*Astrophys. J.* **96** : 28.

THE MASSES OF THE MAGELLANIC CLOUDS FROM RADIO OBSERVATIONS

By F. J. Kerr* and G. de Vaucouleurs†

[*Manuscript received November* 21, 1955]

Summary

The problem of deriving the masses of the Magellanic Clouds is discussed, and provisional estimates are obtained from 21 cm measurements of the distribution of radial velocity.

Various lower and upper limits to the mass of the Large Cloud are found, varying from $1 \cdot 4 \times 10^9$ to $4 \cdot 4 \times 10^9$ solar masses. The best estimate on present evidence is $3 \cdot 0 \times 10^9$. Apart from theoretical considerations, the major uncertainty arises from the sensitivity of the mass to the tilt angle, which is here taken as 65°; the actual value may be somewhat higher and the mass estimate accordingly larger. A provisional value for the mass of the Small Cloud, which can only be estimated by analogy with the Large Cloud, is $1 \cdot 3 \times 10^9$.

With these masses the observed differential velocity indicates that the Clouds probably do not describe a closed orbit around each other.

The mass-luminosity ratio appears to be somewhat lower than corresponding values for other galaxies, supporting the view that the Clouds are young systems.

I. Introduction

This paper is the third of a series based on a survey of 21 cm line radiation from interstellar hydrogen in the Magellanic Clouds. The results of the survey were presented and the brightness observations discussed in paper I (Kerr, Hindman, and Robinson 1954); the motions of the Clouds were then derived from the observed radial velocities in paper II (Kerr and de Vaucouleurs 1955); we now consider the derivation of the masses of the Clouds from the rotational and random motions.

No reliable determinations of the masses of the Clouds are yet available, but rough estimates have been made by Oort (1940), Shapley (1950), and Holmberg (1952). Holmberg's values, which are the most recent, are derived (*a*) from the dispersion of the optical radial velocities measured by Wilson (1917) and (*b*) from an assumed mass-luminosity ratio for a type I system. After correction for the revised distance scale (double the traditional scale), they are of the order of 2×10^9 solar masses for the Large Cloud, and $0 \cdot 6 \times 10^9$ for the Small Cloud. These values are unreliable, however, because (*a*) the optical velocities are few and of low accuracy, and also the main source of the dispersion is not random motions in the Cloud (see paper II), and (*b*) the mass-luminosity

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

† Yale-Columbia Southern Station, Mt. Stromlo, Canberra.

ratio derived from large spirals cannot be safely applied to the Clouds which are so different in their characteristics.

A lower limit to the mass of each Cloud is provided by the mass of neutral hydrogen, which has been deduced from the 21 cm line observations (paper I). These values are $0\cdot6\times10^9$ for the Large Cloud, and $0\cdot4\times10^9$ for the Small Cloud.

In other galaxies, the most satisfactory estimates of mass have been derived from a study of rotational motions. The systems for which the most detailed information has been obtained are the large spirals M31 and M33 (Babcock 1939 ; Mayall and Aller 1942 ; Wyse and Mayall 1942 ; Mayall 1950 ; Baade and Mayall 1951 ; Mayall and Eggen 1953 ; Lohmann 1954 ; Schwarzschild 1954) and the E7 galaxy NGC 3115 (Oort 1940 ; Schwarzschild 1954). In each of these cases the rotational velocity curve has been used to derive the total mass and the mass distribution.

The radio observations of the rotational and random motions are used in the present paper to obtain estimates of the mass of each Cloud and the mass distribution in the Large Cloud. Discussion of the problems arising in the course of this study indicates that more precise values require an extension of the theory, together with the more detailed observations which can now be made.

II. The Mass Derivation Problem

The total mass and the mass distribution can be derived for the Clouds from the rotational velocity curves presented in paper II, following the methods used by previous authors for regular spiral galaxies. The Clouds are, however, more complex dynamical systems than the regular spirals ; their spiral structure is asymmetrical and irregular and the random motions (as indicated by the widths of the line profiles) are comparable with the rotational motions. As far as we are aware, no theory is yet available for such a system.

We will first obtain estimates of the masses of the Clouds by a direct application of the theoretical methods which have been developed for the regular spirals, and then consider the factors which require an extension of the theory. Particular attention will be paid to the case of the Large Cloud, for which the most detailed information is available ; the Small Cloud will be considered only by analogy with the Large Cloud.

The discussion of the rotational motions of the Clouds in paper II was largely based on the median radial velocity of each line profile ; this is the simplest and most objective single parameter for describing a profile whose details and interpretation are not precisely known. After extensive rotation curves had been derived from the median velocities, the rotation could then be considered in more detail. By comparison with the few available optical velocities (Wilson 1917), it was possible to show that the median velocities refer to regions away from the equatorial plane of each system, whereas the peaks of many of the profiles could be associated with that plane. The peak velocities should therefore be used as far as possible in deriving the mass of each Cloud.

A mass estimate obtained from the rotation curve alone will be only a *lower* limit, for at least two different reasons. Firstly, the rotation curve is very insensitive to the mass distributed in the outlying parts of the Cloud, so that the derived value only allows satisfactorily for the mass in the inner regions. We will therefore use other means to extrapolate the mass distribution curve to infinity (Section V (*a*)) ; a reasonable estimate of the required correction can be obtained from the observed distributions of neutral hydrogen and light as a function of radial distance from the centre.

The second shortcoming of a method based on the rotation curve arises from the neglect of the random motions. A mass derivation from a rotational analysis assumes that a system is in dynamic equilibrium under the centripetal force of its own gravitation and the centrifugal force of the rotation. The close connexion between the rotational velocities and the total mass, which is postulated by the theory, probably holds in the large, highly flattened spirals, including our own Galaxy, where the rotational velocity is much larger than the random motions, except near the centre. Even in these cases however the validity of the method has been questioned (Zwicky 1937).

The assumption and theory are much less justified in the case of Magellanic-type galaxies. Apart from the possibility that the Clouds have not yet reached equilibrium, the radio observations have shown that the random and rotational motions are of the same order of magnitude. Such a system is prevented from collapsing by the kinetic energies of both the rotational and random motions, so that both must be taken into account in determining the total mass. Following the treatment based on rotation alone, we will obtain another *lower* limit to the mass from the velocity dispersion alone, neglecting the rotation (Section VI (*a*)).

Finally we will consider the case where the two types of motion are present simultaneously, in order to find closer approximations to the actual mass (Section VI (*b*)).

The main rotational treatment is based on the methods of Wyse and Mayall (1942) and Perek (1948, 1950), but two alternative interpretations are briefly discussed in Section V (*b*).

Theoretical considerations aside, precise values for the masses of the Clouds cannot be expected at this stage, owing to uncertainties in the data, especially in the tilt angle of the Large Cloud ; this angle is taken, in the main discussion, as $i=65°$, but final results will also be given for $i=70$ and 75°. Shapley and Nail (1955) have recently questioned the description of the Large Cloud as a flattened, tilted system (de Vaucouleurs 1955*a*), but the demonstration of rotational motion given in paper II appears to us to confirm that the Cloud is, in fact, a flattened, rotating system, tilted to the line of sight.

For the Small Cloud, uncertainty in the tilt angle is unimportant, but allowance for the asymmetrical " tidal " prominence (de Vaucouleurs 1955*b*) raises a serious difficulty. However, the less extensive velocity curve for the Small Cloud does not permit a full independent treatment, and its mass can only be derived by simple comparison with the case of the Large Cloud.

III. Adopted Rotation Curves

We must first derive the rotation curves which best fit all the available evidence. As shown in paper II, the most appropriate rotation curves are those obtained from the peak radio velocities, which are believed to refer to the equatorial planes of the two systems. On present information, however, such a curve can only be drawn over a limited range ($1° < r < 5°$), but in the case of the Large Cloud we can deduce its probable course in the outer parts with the help of the more extensive median velocity data.

The observed rotation curves are somewhat smoothed, owing to the finite extent of the aerial beam, and we must first attempt to correct for this effect. Bracewell (1955) has given an approximate method for correcting a brightness distribution which has been blurred by a Gaussian aerial beam. In this method a two-dimensional array of values of brightness-temperature is tabulated at intervals of $\sqrt{2}$ times the standard deviation of the Gaussian curve. Then the correction to be applied to the observed value at any point is the difference between that value and the mean of the four surrounding values.

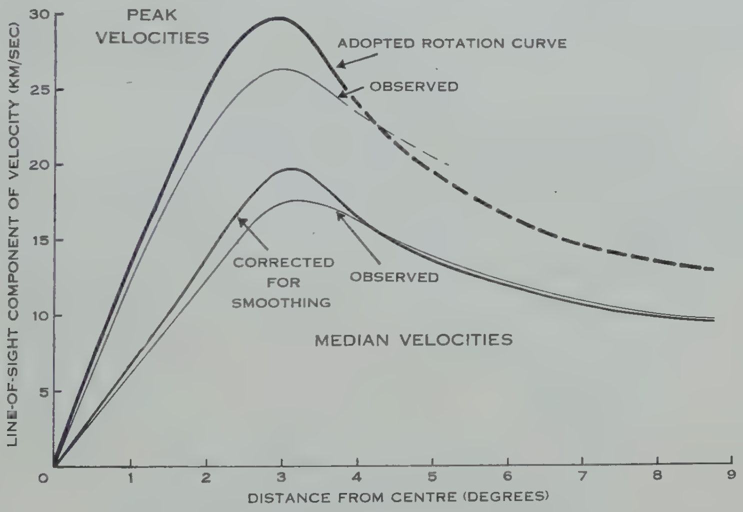


Fig. 1.—Velocity curves for the Large Cloud.

In the present case, the aerial beam does not depart far from a Gaussian shape, with a standard deviation of 0·65°. The method, as described for a brightness distribution, can be applied directly to a velocity distribution if the brightness is uniform over the region considered, but must be modified when, as in the Clouds, the brightness decreases from the centre outwards. This has the effect of giving greater weight to the velocities in the portions of the aerial beam which are nearest the centre of the Cloud. In this case the correction to be applied to an observed velocity can be approximated by weighting the four neighbouring values according to their associated brightness ; this procedure is not formally precise, but gives a sufficiently good approximation under the present circumstances.

This modification of Bracewell's method has been used to correct the rotation curves for the effects of aerial smoothing. The weights were obtained from the mean radial distribution of 21 cm brightness, and idealized velocity patterns over the Clouds were derived from the observed mean rotation curves, assuming tilt angles of 65 and 30° for the Large and Small Cloud respectively. The corrected curves are shown in Figures 1 and 2, together with the observed curves taken (except for the peak velocity curve for the Small Cloud) from Figures 4 and 5 of paper II.

In the case of the Large Cloud, the corrected curve for the peak radio velocities was then extrapolated in the region $4° < r < 8°$, by using the shape of the median velocity curve and assuming that the ratio of peak-to-median velocity tends to unity as $r \rightarrow \infty$.*

The finally adopted curves for the two Clouds are shown by the heavy lines in Figures 1 and 2.

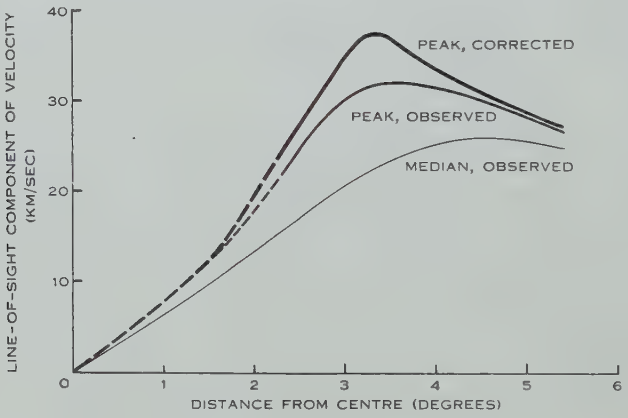


Fig. 2.—Velocity curves for the Small Cloud.

IV. Mass Estimates from Rotation Curve

We now fix attention on the Large Cloud, applying to the rotation curve methods which have already been used for other galaxies.

(*a*) *Point Mass: Keplerian Branch*

A first useful indication of the mass can be obtained from consideration of the Keplerian branch alone. The outer parts of the velocity curve for the Large Cloud agree closely with the curve which would be produced by a central point mass of $1 \cdot 5 \times 10^9$ solar masses.

(*b*) *Wyse and Mayall's Method: Thin Disk Approximation*

In the method developed by Wyse and Mayall (1942) it is assumed that a galaxy can be approximated by an infinitely thin disk having a suitable radially symmetric density distribution. The radial distribution of density in this disk is expressed in the form of a power series extending to the fifth power of the radius and thus involving six parameters. The gravitational force curve corresponding to any density distribution is then derived by a lengthy mathematical process. The force curves produced by a number of special density models are computed and their coefficients tabulated, so that for these particular cases the mass distribution can be readily derived from an observational force curve of the same type. More generally, when none of the special models can be used, five points of a force curve can be used to set up five simultaneous equations, which, together with the condition that the density must be zero outside the disk, yield the density distribution.

---

* The trend of the ratio towards unity was found to be approximately linear with $1/\sqrt{r}$ over the region where both velocities were available; this trend was continued in the extrapolation.

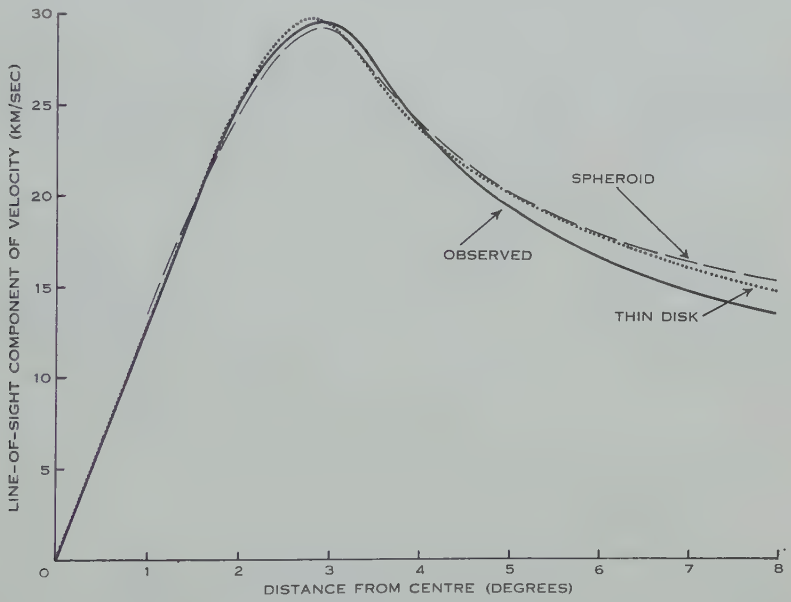


Fig. 3.—Theoretical velocity curves for the Large Cloud for the best fitting thin disk and spheroid models, compared with the adopted observed curve.

An application of the full treatment to the observed velocity curve (Fig. 3) has led to unsatisfactory results for the inferred density distribution, mostly on account of the alternating character of the series involved ; the desired quantities are obtained as the differences between much larger coefficients of opposite signs. This defect had already been noted by Wyse and Mayall.

In view of this a simpler treatment was preferred, based on the theoretical distributions for some of Wyse and Mayall's thin disk models. It was found by trial and error that two of their models could be combined to give a force curve in fair agreement with the observed one. The models used were those with an " oblate spheroid " density distribution, $\sigma=\sigma_c[1-(r/R)^2]^{\frac{1}{2}}$ and a " Gaussian " distribution $\sigma=\sigma_c \exp\{-4(r/R)^2\}$. The best fit was obtained with a composite model, consisting of three spheroids with $R=2\cdot2$, $2\cdot9$, $3\cdot5°$, and $\sigma_c=33$, 67, 33 solar units per square parsec, together with a " Gaussian " distribution with $R=5\cdot37°$ and $\sigma_c=22$. The theoretical velocity curve for this model is shown in Figure 3, in comparison with the observed curve, and the distribution of projected density in Figure 4. The model leads to a mass of $1\cdot73\times10^9$ solar masses for the Large Cloud and a projected central density of 155 suns/pc^2.

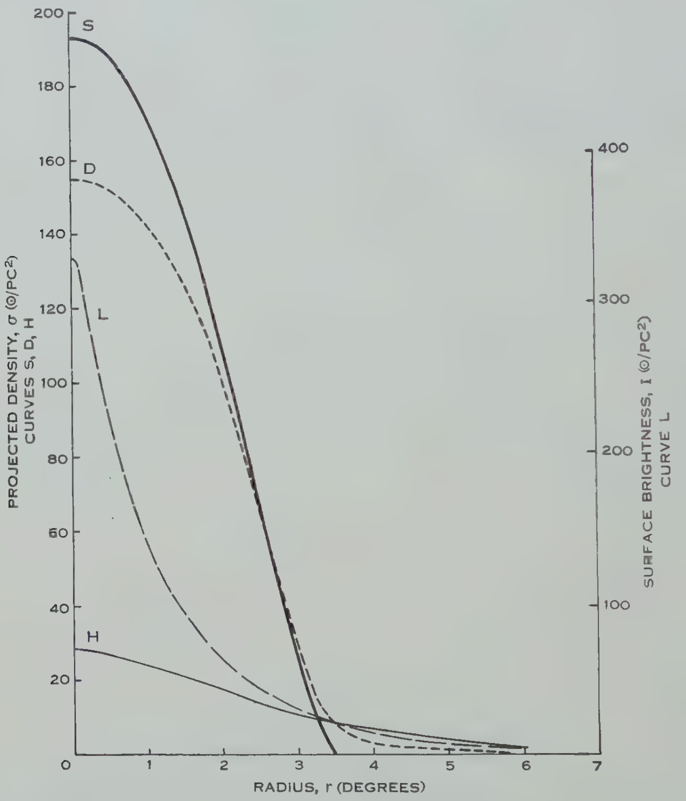


Fig. 4.—Projected density distribution in the Large Cloud, for H I, red light, and rotational mass according to the two models :

——— Hydrogen, corrected for aerial smoothing $(\odot/\mathrm{pc}^2)$;
— — Red light $(\odot/\mathrm{pc}^2$, right-hand scale) ;
- - - - Thin disk $(\odot/\mathrm{pc}^2)$;
——— Spheroid $(\odot/\mathrm{pc}^2)$.

(c) *Perek's Method : Oblate Spheroid Approximation*

The thin disk treatment has given a good fit with the observational results, but the method has the disadvantage that it neglects the thickness of the system and thus prevents the derivation of space densities. In the case of NGC 55 (Plate 3, paper II), an edgewise system believed to be of Magellanic type, it can be seen that the thickness at right angles to the equatorial plane is by no means negligible ; a microphotometric investigation (de Vaucouleurs 1956) indicates that the ratio c/a of the axes of the outer isophotes of NGC 55 is close to 1/5. We will adopt this value for the Large Cloud.

A better approximation may then be obtained by following Perek (1948, 1950) who has discussed theoretically the velocity curves of non-homogeneous spheroids for various density laws and produced a convenient nomogram and table for the rapid derivation of the force curves as a function of c/a, when the density law is of the form

$$\rho=\rho_c(1-m^2)^n,$$

where

$$m^2=(x^2+y^2)/a^2+z^2/c^2.$$

In the present case a simple model with $n=1$ was found to give a good fit with the observed velocity curve (see Fig. 3) for

$$a=3\cdot 5^\circ=2\cdot 8\text{ kpc},$$
$$\rho_c=1\cdot 77\times 10^{-23}\text{ g/cm}^3=0\cdot 26\text{ sun/pc}^3.$$

The corresponding values for the projected central density σ_c and total mass $\mathbf{M}$ are

$$\sigma_c=0\cdot 266a\rho_c=195\text{ suns/pc}^2,$$
$$\mathbf{M}=0\cdot 335a^3\rho_c=1\cdot 93\times 10^9\text{ solar masses.}$$

The total mass is in good agreement with the value derived from the thin disk model. The projected central density σ_c is slightly higher, as could be expected from the change to a model with matter outside the equatorial plane, but we obtain in addition the central space density ρ_c, which is equivalent to 11 hydrogen atoms per cubic centimetre (either condensed into stars or in interstellar space).

A defect of the model is that it leaves no room for matter beyond $r=3\cdot 5^\circ$ where both radio and optical observations detect considerable radiation. Experiments with other values of c/a indicate that non-zero densities in the outer parts can be obtained if c/a is less than 1/5 and the thin disk approximation gives then a fair idea of the sort of distributions which can be obtained. For $c/a=1/5$, matter can be spread beyond $r=3\cdot 5^\circ$ by using more elaborate density laws, but the accuracy of the present data was not considered sufficient to justify this refined treatment. In any case the total mass and central projected density cannot be changed much and the values above should be sufficient for the present, since more serious sources of uncertainty exist, as described earlier.

For the succeeding discussion, we will adopt $1\cdot 8_5\times 10^9$ solar masses for the mass of the Large Cloud indicated by the rotation curve, for a tilt angle $i=65^\circ$.

V. Discussion of Mass Estimates Derived from Rotation

(a) *Extrapolation to Infinity*

The mass derived from the radio rotation curve is probably only a lower limit, since the velocity curve is insensitive to the density in the outlying regions. This shortcoming has already been experienced by Wyse and Mayall (1942) in their discussion of M33. The density in these outer regions is low, but the total volume of space concerned is large, so that the contribution to the mass might be quite substantial.

Evidence that the thin disk and spheroid models provide too little mass in the outer parts of the Large Cloud, in spite of the good fit which they give to the observed velocity curve, is shown in Figure 4. The mean radial distributions of projected mass corresponding to these models are compared with those for neutral hydrogen, and for the bulk of the stellar population as indicated by observations of the surface brightness in red light (unpublished data). Both the latter distributions indicate that there is more mass in the outlying parts than has been accounted for by the models ; the divergence becomes serious at about 3°.

Our problem then is to form some estimate of the factor by which the mass of the models must be increased to allow for the mass distributed in the outer parts. A lower limit to this additional mass can be obtained directly by integrating the hydrogen density in excess of the density given by the models. Another estimate may be obtained by assuming that the distribution of the combined mass (stellar and interstellar) in the outer parts is intermediate between the hydrogen and light distributions. This assumption is plausible, since the hydrogen is known to comprise by far the largest fraction of the interstellar mass, and the faint stars which contribute most of the red light carry most of the stellar mass.

TABLE 1

OVERALL MASSES OF NEUTRAL HYDROGEN

	LMC	SMC
Measured	$0\cdot57\times10^9$*	$0\cdot43\times10^9$†
Extrapolated	$0\cdot7{:}\times10^9$	$0\cdot5{:}\times10^9$

* $A=165$ sq. deg. ($\bar{r}=7\cdot3°$).
† $A=104$ sq. deg. ($\bar{r}=5\cdot7°$).

Before these estimates can be formed, however, we note that the hydrogen density, as indicated by the trend of the curve in Figure 4, does not fall abruptly to zero beyond the last measurable contour. An attempt has been made to allow for the mass of hydrogen beyond the observed limits by an extrapolation method similar to that used in the determination of the total (integrated) luminosities in galaxies (Redman 1936 ; de Vaucouleurs 1948).

Graphical integration of the H I brightness contours (paper I) gives the total mass of hydrogen within successive contours ; if the logarithm of this mass is plotted as a function of the area A limited by the contour (e.g. against $1/\sqrt{A}$), it is possible to obtain by a short extrapolation the limit towards which the mass of hydrogen tends as A is indefinitely increased. In this way allowance may be made for the contribution of the very faint outer regions where the radiation falls below the threshold of detectability.

Extrapolation of the measured masses of neutral hydrogen leads to the rather conservative estimates of the overall masses which are shown in Table 1.

Once in possession of the overall mass the fraction of the total within any distance from the centre can be readily computed, leading to the relative integrated mass curves shown in Figure 5.

The same procedure has been applied to an unpublished set of isophotes in red light extending to an average radius $\bar{r}=7\cdot7°$ ($A=185$ sq. deg.) and including about 85 per cent. of the total (integrated) luminosity (see Fig. 5). Note, however, that the centre of these isophotes is displaced from the centre of rotation or the hydrogen centre.

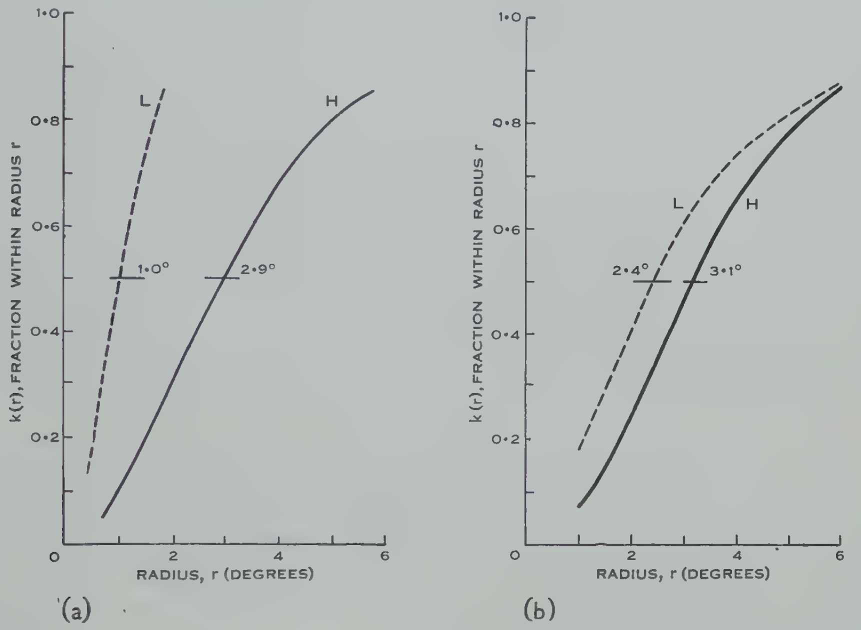


Fig. 5.—Relative integrated intensity distributions in the Clouds for neutral hydrogen and red light. The curves give the fraction $k(r)$ projected within a circle of radius r centred on C_H or C'. The effective radii r_e are shown on the curves at $k(r)=0\cdot5$. (*a*) Small Cloud, (*b*) Large Cloud.

We can now estimate the mass of hydrogen which has not been accounted for by the models. In this way, we obtain $0\cdot3\times10^9$ solar masses as a lower limit to the excess mass. The contribution from the stars cannot be directly estimated in a similar manner, because the optical surface brightness is probably not proportional to the projected surface density of stellar mass (cf. Section V (*b*) below).

As a first approximation, however, we can assume that the combined mass (gas+stars) is distributed in the same way as the mean of the hydrogen and light, and that the value of the mass derived from the rotational analysis is unaffected by the mass of the outlying regions. The curves of Figure 4 indicate that the total mass of hydrogen is $2\cdot1$ times that within $r=3°$ from the centre; the corresponding factor for the luminosity is $1\cdot7$. For an estimate of the mass we will take a factor of $1\cdot9$.

The mass within $r=3°$ indicated by the rotational analysis is $1 \cdot 6 \times 10^9$ for the thin disk model and $1 \cdot 8 \times 10^9$ for the spheroid model. The corrected mass, extrapolated to infinity according to the above assumption, is therefore $1 \cdot 9 \times 1 \cdot 7 \times 10^9 = 3 \cdot 2 \times 10^9$ solar masses.

This estimate has been tested by comparing a theoretical velocity curve for the extrapolated mass distribution with the observed curve. The theoretical curve was derived by approximating the mass distribution by a series of straight-line segments, and then applying Wyse and Mayall's solution for a linear distribution of projected density. The velocities obtained in this way were considerably higher than the observed velocities in the outer parts of the Cloud, e.g. 27 and 26 km/sec at $r=5°$ and 7°, where the observed values were $19 \cdot 7$ and $13 \cdot 7$ km/sec. The observed velocities are of low accuracy in the outer regions, but the above differences are substantially greater than the estimated probable error.

The above estimate of the extrapolated mass must therefore be too high. There are two possible reasons for this, one or both of which may be operating: either the assumption of a nearly constant mass-luminosity ratio is at fault (see also the discussion of Schwarzschild's method in Section V (*b*)), or else the shortcomings of the simple rotational analysis may be connected with the neglect of the random motions (see Section VI).

For the present, we can only conclude that the extrapolated mass is $>2 \cdot 1$ and $<3 \cdot 2 \times 10^9$, say $2 \cdot 5 \times 10^9$ solar masses.

(*b*) *Other Interpretations of the Velocity Curve*

Our mass estimates have been obtained through two related interpretations of the velocity curve (thin disk and spheroid models), based on some *assumed* density distribution laws selected for simplicity. The rather good fit they give to the observed velocity curve does not necessarily imply that the actual density distribution closely follows the assumed laws everywhere. In fact Figure 4 shows conclusively that it does not, at least in the outer parts.

It is therefore desirable to consider two alternative interpretations of the velocity curve which have been proposed for other galaxies:

(1) that the force curve can be accurately represented by the simple interpolation formula due to Bottlinger (Lohmann 1954);

(2) that the projected surface brightness is proportional to the projected mass density, i.e. that the mass-luminosity ratio is constant throughout the system (Schwarzschild 1954).

According to Lohmann, following Bottlinger, the gravitational force in the equatorial plane of a galaxy can be represented by

$$F = ar/(1+br^3),$$

and

$$v^2 = ar^2/(1+br^3),$$

whence

$$\mathbf{M} = a/bG \quad (G = \text{gravitational constant}).$$

A model for which $v_{\text{max.}} = 30 \times \sec 65° = 71$ km/sec at $r = 3° = 2 \cdot 4$ kpc ($a = 2 \cdot 67 \times 10^3$ (km/sec)^2(kpc)$^{-2}$, $b = 0 \cdot 147$ (kpc)$^{-3}$) gives a total mass

$\mathbf{M}=4\cdot2\times10^9$ solar masses, but as shown in Figure 6 it does not give a good fit with the observed velocity curve. If the theoretical curve is adjusted to fit at 2 or 4°, the corresponding masses are $3\cdot5$ and $2\cdot9\times10^9$.

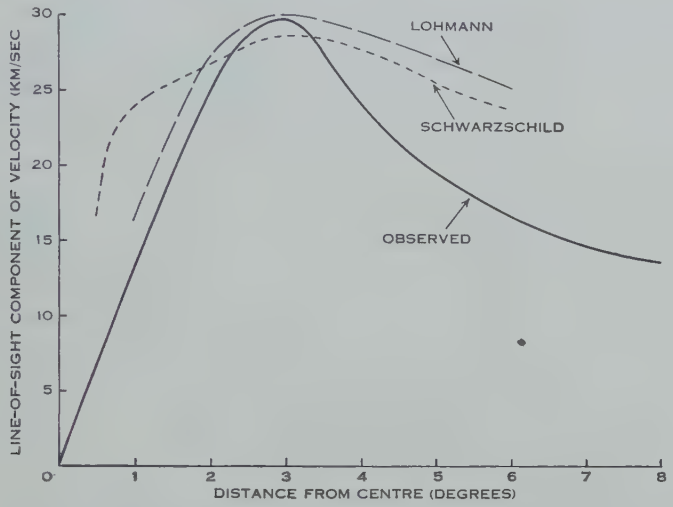


Fig. 6.—Comparison between theoretical velocity curves derived according to Lohmann's and Schwarzschild's assumptions and the observed velocity curve.

Schwarzschild assumes that the projected mass distribution $\sigma(r)$ is proportional to the luminosity distribution $I(r)$; the latter is approximated by a series of straight segments, so that

$$I(r)=\sum_n a_n(1-r/R_n) \quad \text{and} \quad \sigma(r)=\sum_n A_n(1-r/R_n),$$

TABLE 2

STRAIGHT SEGMENT APPROXIMATION TO LMC LUMINOSITY CURVE

n	R_n (pc)	a_n ($\odot$/pc²)	$R_n^2 a_n$
1	400	40	6×10^6
2	800	80	51×10^6
3	1800	140	455×10^6
4	3200	53	510×10^6
5	4800	18	440×10^6
6	7200	9	465×10^6

where $A_n=f\times a_n$, if f is the supposedly constant mass-luminosity ratio. Wyse and Mayall's results for a linear density distribution then lead to a simple solution of the velocity curve for the thin disk model, and to a total mass

$$\mathbf{M}=\tfrac{1}{3}\pi\sum_n A_nR_n^2 \quad \text{(in solar units).}$$

This treatment was applied to an unpublished luminosity curve in red light for which the adopted straight segment approximation is given in Table 2.

Allowing for various numerical factors this leads to a total mass $\mathbf{M}=2\cdot 0\times 10^9\times f$ (in solar units) and a computed velocity curve

$$v(r)=0\cdot 131\sqrt{f}[r\sum_n a_n g(r/R_n)]^{\frac{1}{2}},$$

where $g(r/R_n)$ is the function $M(\alpha)$ and $m(\beta)$ tabulated by Wyse and Mayall. Approximate agreement with the observed maximum of the velocity curve is obtained for $f=1\cdot 5$ (if $i=65°$), so that $\mathbf{M}=3\cdot 0\times 10^9$ solar masses, but elsewhere the fit is very poor (Fig. 6).* The best average fit over a larger portion of the velocity curve would give $\mathbf{M}=2\times 10^9$.

These two alternative interpretations which take some account of the mass in the outlying parts are useful in confirming the order of magnitude of the derived mass ; we see from the velocity curves, however, that their assumptions do not fit the observations as closely as do those of the earlier models, and we will base our conclusions on the results obtained in Sections IV and V (*a*).

VI. Mass Estimates including Random Motions

(*a*) *Spherical Approximation*

As mentioned in Section II above, the random, or peculiar, motions in the Large Cloud appear to be comparable with the rotational motions. This suggests that another lower limit to the mass of the Cloud may be obtained from the velocity dispersion through the virial theorem, as applied to spherically symmetrical systems in which rotation is negligible (spheroidal galaxies, globular clusters). In a sense this represents the limiting case at the opposite extreme from the thin disk approximation in the rotational analysis ; the actual system is in an intermediate situation.

The mean value of the observed velocity dispersion (the half half-width of the line) in the Large Cloud is about 25 km/sec, with slightly higher values in the central regions and slightly lower values in the outer parts. This velocity spread must be due in part to systematic mass motions in the equatorial plane, as shown by the irregularity of many of the profiles, and also to the variation of rotational speed with depth in the Cloud. There is also a small broadening due to the finite receiver bandwidth. Hence the true " random " velocities must be somewhat smaller than 25 km/sec. The various effects cannot be separated on present evidence, but for definiteness we will take a value of 20 km/sec for the r.m.s. random velocity.

For a numerical application of the virial theorem, we use the simple equation (Chandrasekhar 1942, p. 200)

$$\mathbf{M}=\frac{2\overline{V^2}R}{G},$$

where $\overline{V^2}=\overline{u^2}+\overline{v^2}+\overline{w^2}=3\overline{w^2}$, R is a suitably defined effective radius, and G is the gravitational constant. Here we may take $(\overline{w^2})^{\frac{1}{2}}=20$ km/sec and $R=3°=2\cdot 4$ kpc, which gives $\mathbf{M}=1\cdot 4\times 10^9$ solar masses.

---

* Reference to Figure 3 in Schwarzschild's paper indicates that his attempt to fit the M33 data led to the same sort of discrepancy. It seems clear that the assumption $f=$const. is not justified for late-type systems.

This value must be greater than the mass which can be related to the random motions, since the Cloud is far from spherical, but it must be taken as a lower limit, as the rotational motions have been neglected.

(b) *Combined Effect of Rotational and Random Motions*

Direct addition of the masses obtained by considering rotational and random motions separately leads to a value of

$$2 \cdot 5 + 1 \cdot 4 = 3 \cdot 9 \times 10^9 \text{ solar masses.}$$

This can be regarded as an extreme upper limit, as the effects are obviously not additive and their interaction must be taken into account. We will estimate the combined centrifugal effect of the two types of motion from an approximate physical picture.

When random motions are superposed on the observed rotation, each particle will execute an orbit which is in general elliptical and inclined to the equatorial plane. The components of the random velocities which are directed towards the centre (the x-direction) will average to zero in their centrifugal effect ; they can therefore be neglected in this discussion, and the orbits regarded as circular. The velocity of a particle " rotating " in such an inclined orbit will in general be greater than the equatorial plane component of rotational velocity which was derived in paper II. The higher rotational velocities can support, by their centrifugal action, a greater mass.

To a first order, the representative velocity, which is to be associated with the mass of the system, can be taken as that derived from the equatorial plane rotational velocity and the r.m.s. value of the component of random velocity directed away from the plane, the z-component. The x-component has no net centrifugal effect, as explained above : the y-(tangential) component will not to the first order affect the result, since the mean rotational velocity as measured is a fair approximation to the vector sum of the pure rotational velocity and the random y-component.

We again take 20 km/sec as the r.m.s. random velocity in the line of sight for the Large Cloud. In a spherical system, the three components of the random motions would be equal, and equal to the line-of-sight component. Since the Cloud appears to be flattened, the z-component of the random velocities is presumably somewhat smaller than the x- and y-components, but for the present approximate treatment we will regard the three components as equal to 20 km/sec. Further, the observations suggest that the random velocities decrease on going outwards from the centre of the Cloud ; they also presumably decrease on going away from the equatorial plane, but for the moment we will neglect both these variations.

The maximum rotational velocity in the equatorial plane curve is 71 km/sec (for a tilt angle of 65°), but we should take here a mass-weighted mean value for the rotational velocity, say 45 km/sec. Compounding this with the random z-component of 20 km/sec gives a correcting factor of about $1 \cdot 2$. The previous value of $1 \cdot 8 \times 10^9$ solar masses, without the incompleteness correction, is thereby increased to $2 \cdot 2 \times 10^9$. If, finally, we apply the incompleteness correction, we get a value for the total mass of $3 \cdot 0 \times 10^9$ solar masses, with limits of $2 \cdot 2 + 0 \cdot 3 = 2 \cdot 5 \times 10^9$ and $2 \cdot 2 \times 1 \cdot 9 = 4 \cdot 2 \times 10^9$.

Oort (1940), following Jeans (1922), has given a treatment for a system in a dynamically steady state. He obtained a relation between the " circular " velocity in the absence of random motions, V_c, and the observed mean rotational velocity when random motions are present, V_0, which may be written, in our notation,

$$V_c^2 - V_0^2 = \overline{u^2}\left[-\frac{\partial \log \rho}{\partial \log r} - \frac{\partial \log \overline{u^2}}{\partial \log r} - \left\{ 1 - \frac{\overline{v^2}}{\overline{u^2}} \right\} \right],$$

where u,v,w are the x,y,z components of the random velocity, and ρ the space density.

The assumption of strict equilibrium is not applicable to the Large Cloud, and also the expression requires a greater knowledge of the random velocities through the system than is at present available, but an attempt to estimate the factors involved leads to a mean value of V_c^2/V_0^2 in rough agreement with the correction factor derived above.

VII. Best Estimate of the Masses

(*a*) *Large Cloud*

In Table 3 are collected the various upper and lower limits for the mass of the Large Cloud which have been derived in the preceding sections. The rotational treatment leads in the first instance to a mass of $1 \cdot 8 \times 10^9$ solar masses, and $2 \cdot 2 \times 10^9$ when allowance is made for the random motions. If the estimated extrapolation correction is applied to the latter value, a mass of $3 \cdot 0 \times 10^9$ is obtained. This is the best estimate of the mass of the Large Cloud which can be made on the present evidence, and subject to the various assumptions which have been introduced. As pointed out in Section V (*a*), however, the extrapolation correction is very uncertain, although it is clear that such a correction must be applied. Other treatments of the data which make some

Table 3

ESTIMATES OF THE MASS OF THE LARGE CLOUD

(in 10^9 ☉, for $i = 65°$)

Method	Inner Parts ($r < 3°$)	Main Body (models)	Whole Cloud (extrapolated)
Rotation only	$\geqslant 1 \cdot 7$	$\geqslant 1 \cdot 8$	$> 2 \cdot 5$
Random motions only	$\geqslant 1 \cdot 4$		
Rotation + random ..		$> 2 \cdot 2$	$3 \cdot 0$ $(\leqslant 4 \cdot 2)$
Lohmann			(3–4)
Schwarzschild			(2–3)
Best estimate			$3 \cdot 0$

allowance for the outlying mass (Section V (*b*)) give values of the same order, 3×10^9 solar masses.

The greatest uncertainty in the data lies in the tilt angle of the Large Cloud ; the whole analysis has been carried out with a tilt angle of 65°, the provisional value derived from the ellipticity of the outer optical isophotes (de Vaucouleurs 1955*a*), but the radio data suggest that the tilt may well be greater than this. Since the square of the secant varies rapidly in this range of angle, the consequent uncertainty in the mass is very large. For tilt angles of 70 and 75°, the corresponding values for the mass would be $4 \cdot 3 \times 10^9$ and $7 \cdot 1 \times 10^9$ solar masses respectively.

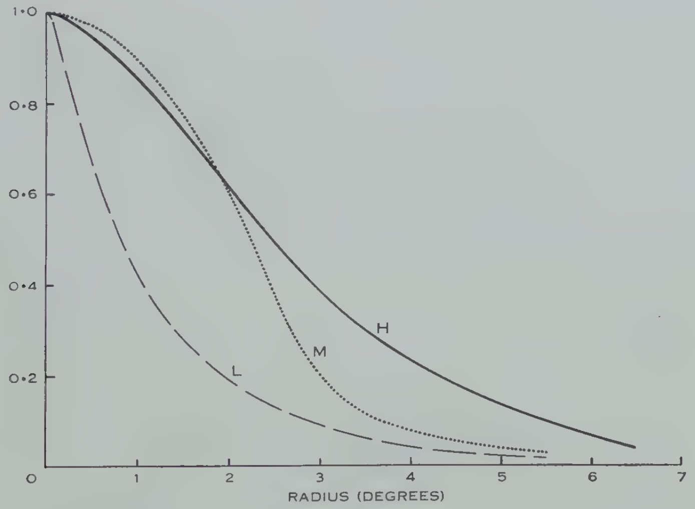


Fig. 7.—Comparison between the approximate mass distribution in the Large Cloud, derived from the present treatment, and the observed distributions for hydrogen and red light.

The mass distribution cannot be determined in the outer parts, but a distribution curve which is consistent with the above estimate for the extrapolated mass is shown in normalized form in Figure 7, together with the hydrogen and luminosity distributions.

(*b*) *Small Cloud*

We now discuss briefly the Small Cloud. The rotation curve in this case does not extend sufficiently far from the centre for an independent treatment to be possible, but an approximate mass can be obtained from the Large Cloud discussion if we assume that both Clouds are basically built on the same model. The optical observations of spiral structure give some support to this assumption but the presence of the asymmetrical prominence and the displacement of the centre of rotation from the optical centre detract from its validity.

For a given model, $\mathbf{M} \propto (v_m)^2 \times r_m$, where v_m is the maximum velocity, reached at a distance r_m from the centre. From the adopted rotation curve for the Small Cloud (Fig. 2), v_m=37 km/sec at r_m=3·3°, allowing for a tilt angle of 30°. If similar models are taken, the mass is therefore 45 per cent. of that for the Large Cloud, when the Large Cloud tilt is taken as 65°. This leads to a value for the mass of the Small Cloud of $1 \cdot 3 \times 10^9$ solar masses. Although the value was arrived at through a comparison with the Large Cloud, it is not itself affected by the uncertainty in the tilt angle for the Large Cloud, and the effect of the uncertainty in the tilt of the Small Cloud amounts to only ±10 per cent.

A ratio of 45 per cent. between the masses of the two Clouds appears to be somewhat high, since the total luminosity of the Small Cloud is only a fifth of that of the Large Cloud, according to the most recent determinations; also, if the Small Cloud prominence is due to tidal perturbation by the Large Cloud, the mass of the Small Cloud must be considerably lower than that of the Large Cloud. The structures of the two Clouds may well be sufficiently different for the mass of the Small Cloud to have been overestimated by the above procedure. Also, the discrepancy would be reduced by using a larger tilt angle for the Large Cloud, as shown in Table 4, which also gives the results of comparisons between the two Clouds from other methods of observation.

TABLE 4

COMPARISONS BETWEEN THE SMALL AND LARGE CLOUDS

Quantity	SMC/LMC
Mass (assumed i for LMC=65°)	0·45
(,, ,, ,, =70°)	0·31
(,, ,, ,, =75°)	0·19
Neutral hydrogen (including SMC prominence)	0·71
,, ,, (excluding SMC prominence)	0·33
Luminosity	0·19

The Small Cloud presents a special problem through its large prominence, which is of much greater relative importance in the radio than in the optical view. In fact, the greater part of the gas appears to be outside the main body of the Cloud as defined by the distribution of stars.

An attempt has been made to obtain separate values for the mass of hydrogen in the prominence and in the main stellar system. The results are necessarily rough since the two parts are not clearly distinguishable. In particular, optical studies indicate that the prominence is not in the equatorial plane of the flattened stellar system (de Vaucouleurs 1955*b*).

For this purpose, the main stellar system was taken to be limited by an ellipse (dashed contour on Plate 2 of paper II) centred on the optical centre, C, with major axis 9° and minor axis 4°, the approximate " overall " dimensions indicated by the star counts at $m=16$ (de Vaucouleurs 1955*b*). Integration over the hydrogen contours gave $0 \cdot 2 \times 10^9$ solar masses for the mass of hydrogen in the main stellar system, and $0 \cdot 22 \times 10^9$ in the prominence. Star counts indicate that 70 per cent. of the stars brighter than $m=14 \cdot 3$ are in the main system, and only 30 per cent. in the prominence. Counts to $m=16$ do not yet allow a similar comparison, but it is clear that the percentage of the total population located within the main stellar system is even higher. Thus the ratio of gas to stars is very different in the two parts of the Small Cloud. The great difference in the distributions of gas and stars in the Small Cloud is evident from Figure 5 ; in the Large Cloud, the distributions are much more similar.

VIII. Orbital Motion of the Clouds

It was shown in paper II that the differential radial velocity of the Clouds is about 50–60 km/sec, the exact value depending on the galactic rotation velocity at the Sun. If the two Clouds were moving as an isolated system, their combined mass $\mathbf{M}$ would be related to the relative orbital velocity v by the expression

$$\mathbf{M}=\frac{v^2}{G(2/r-1/a)},$$

where G is the gravitational constant, r the distance between the two Clouds ($\geqslant$16 kpc, the projected distance), and a the semi-major axis of the relative orbit. The minimum value of $\mathbf{M}$ for a closed orbit would correspond to a parabolic orbit ($a=\infty$), seen edgewise, and with the system at periastron. The observed value of differential radial velocity could represent motion in a closed orbit only if the sum of the masses of the Clouds exceeded 5×10^9 solar masses. For a circular orbit, the combined mass would be greater than 10×10^9 solar masses. The sum of the masses derived above is less than 5×10^9 solar masses, but would exceed this value if the tilt angle of the Large Cloud is increased from 65° to 70° or greater. The total mass is, however, hardly likely to approach 10×10^9. We must therefore conclude that the Clouds are moving in a hyperbolic or nearly parabolic orbit relative to one another.

The Clouds cannot in fact be considered independently of the Galaxy, since the galactic field at the Clouds is of the same order of magnitude as that of either Cloud at the other. The group can only be treated as a three-body system. The evidence that the Clouds are probably moving in an open orbit carries a broader inference in suggesting that caution must be adopted in any study in which the combined masses of double galaxies are derived from their relative motions, particularly when the galaxies are located in a cluster.

IX. Comparisons with Other Galaxies

As noted in Section I, only two other external galaxies, M31 and M33, both regular spirals, have been previously investigated for rotation and mass distribution in anything like the detail given by the radio observations in the Magellanic Clouds. To these may be added the still fragmentary information available on our own Galaxy. It is therefore important to compare the new results with these earlier data to see whether they throw additional light on the dynamics and evolution of stellar systems.

The theoretical difficulties discussed in earlier sections all apply, to a greater or less extent, to the regular spirals considered by other authors. Values derived on a similar basis must therefore be selected in making any comparison between the mass of either Cloud and the mass of some other galaxy. No correction for " incompleteness " or for random motions has been made in the treatments used for other galaxies ; the latter correction will, however, be quite small for the regular spirals, so need not be considered here.

An illustration of the effect of the various treatments on mass estimates may be obtained from a comparison of some values computed for M33 from the same set of optical data (Mayall and Aller 1942). The data are collected in Table 5 ; Schwarzschild's and Lohmann's values are as published ; Wyse and Mayall's values from thin disk models are corrected for the doubling of the distance scale. The value corresponding to the present treatment of the Large Cloud was obtained as follows. Since spheroid models give values 10–20 per cent. higher than thin disk models (cf. Section IV), we may take $4\cdot0\times10^9$ as our corresponding estimate of the mass within $r=18'$. From the observed luminosity distribution in M33 (Patterson 1940), we compute that about 75 per cent. of its total luminosity is comprised within this same radius ; hence the total mass extrapolated as explained in Section V (*a*) is about $5\cdot5\times10^9$.

TABLE 5

MASS ESTIMATES FOR M33

Wyse and Mayall (1942) (corrected) $r<18'$ ..	$3\cdot5\times10^9$
$r<30'$ ..	$4\cdot0\times10^9$
Present treatment	$5\cdot5\times10^9$
Schwarzschild (1954)	5×10^9
Lohmann (1954)	10×10^9

Various comparisons between the Clouds and other galaxies are summarized in Table 6.

Consider first the periods of rotation. Assuming circular motions the period of rotation at any point is

$$P=2\pi r/v_r=6\cdot16\times10^9/\omega \qquad \text{(years)}$$

if the angular velocity ω is expressed in km sec^{-1} kpc^{-1}. The values of P listed in Table 6 correspond to the maximum of the velocity curve, except for the last two cases. The periods found for the Clouds, 350 and ~150 million years, are as long or longer than those in the corresponding parts of regular spirals and elliptical galaxies. According to some provisional data of Mayall (1948) the periods of rotation in the nuclei and " main bodies " of regular galaxies (corrected for the revision of the distance scale) are as follows :

E7, S0	5–10 million years
Sa	10–20 million years
Sb, SBb ..	10–40 million years
Sc	20–80 million years.

There is, however, considerable scatter and some late Sc and SBc galaxies indicate periods well in excess of 100 million years.

With the small sample of galaxies for which reliable mass estimates are available at present, the probable correlations between period of rotation, size,

TABLE 6

ELEMENTS OF SOME GALAXIES

Object	Type	Diameter (" main body ") (kpc)	Absolute Magnitude	Luminosity **L** (10^9 $\odot$)	Period of Rotation(1) P (10^6 years)	Mass **M** (" main body ") (10^9 $\odot$)	**M/L**	Projected Central Density ($\odot$/pc^2)
SMC(2)	Mpec	4	$-16\frac{1}{2}$	0·5	350	$\sim$1	$\sim$2	?
LMC(2)	M	8	$-18\frac{1}{2}$	2·5	200 to 100(2)	2·5 to 5(2)	1 to 2(2)	200 to 400(2)
M33	Sc	8	−18	2·0	100	5	2 to 3	500 ?
Galaxy	Sbc ?	20	?	?	185(3)	100	(4)(4)	?
M31	Sb	20	−20	12	200	150	12	2000 ?
NGC 4594(5) ..	Sa	5:	−19:	5:	40:(6)	>60:	>12:	?
NGC 3115(5) ..	E7	5:	−18:	2:	10:(6)	>40:	>20(5)	>4000 ?

References : Oort (1940, Table 4) ; Mayall and Aller (1942, pp. 20–21) ; de Vaucouleurs (1953, Table IV) ; Lohmann (1954, Table 3) ; Schwarzschild (1954, Table VII).

(1) At $r=r_m$.

(2) For range of i from 65 to 75°.

(3) From 21 cm rotation curve (Kwee, Muller, and Westerhout 1954).

(4) Near the Sun.

(5) The masses of NGC 4594 and 3115, being " rotational " masses, are lower limits only. Their distances, which are highly uncertain, have been taken here as 4 Mpc in the revised scale. The value **M/L**$=100$, obtained by Schwarzschild (1954) for NGC 3115, is for 2 Mpc, corresponding to 50 at 4 Mpc.

(6) In these cases the periods of rotation refer to the inner regions showing solid body rotation. (Detailed velocity curves have not been published.)

and mass cannot be eliminated, but existing information is consistent with the assumption that spin increases along the sequence (M)→(S)→(E); this may therefore represent an evolutionary sequence as suggested by Shapley (1950). Further, the situation in Magellanic-type systems is compatible with some views put forward by Hoyle (1951) and von Weizsäcker (1951) on the origin of rotation in galaxies: this is indicated by the slow rotation and large random motions which result in irregularity of the spiral pattern, together with the frequent occurrence of such systems in close pairs showing clear signs of strong interaction (de Vaucouleurs 1954). If this is substantially correct, then the Magellanic Clouds appear to be galaxies of recent formation and still in the process of condensation and organization; thus the dynamical evidence supports the interpretation suggested by the detailed study of the physical content of the Clouds, especially of the Large Cloud, which shows large numbers of superluminous and presumably "young" stars of short lifetime.

Additional evidence in support of the interpretation that the Clouds are young systems is found in the low values obtained for the mass-luminosity ratio and central densities as compared with the regular spirals or the solar neighbourhood (Table 6). In combination with other optical data this suggests that stars fainter than the Sun, for which the mass-luminosity ratio exceeds unity, are relatively rare in the Clouds. According to current ideas (Strömgren 1952), such low luminosity stars have very long lifetimes and are not likely to condense out of interstellar matter under present conditions; the near absence of such "old" stars in the Clouds adds colour to the hypothesis that such systems are of comparatively recent formation.

The 21 cm results have provided estimates of both the mass of neutral hydrogen and the total mass. For the Large Cloud, assuming a tilt angle $i=65°$, the ratio of these quantities is 23 per cent. for the extrapolated values, or 19 per cent. when the more reliable results for $r<3°$ are considered. (The ratio will be smaller for $i>65°$.) These values may be compared with the corresponding figures for the relative gas density in the solar neighbourhood, for which two recent estimates are 15 per cent. (Oort 1952) and 20 per cent. (Bok 1954).

X. CONCLUSION

The survey on which this study was based was of a preliminary nature, but, even at the present stage, the velocity curve for the Large Cloud is more extensive than, and probably as accurate as, that available for any other external galaxy. A reasonable estimate of the total mass of the Large Cloud has been obtained, but the evidence on mass distribution is weak, as is shown by the discussion on extrapolation to infinity of the results for the inner regions. Also, little is yet known about the detailed internal motions of the Clouds.

The shortcomings in each case are partly observational and partly theoretical. The observational requirement is for greater resolution, both in angle and velocity, in the main body of each Cloud, and for higher sensitivity in the outer regions; on the theoretical side, a better treatment is needed for a thick system containing both rotational and random motions.

XI. ACKNOWLEDGMENT

The authors wish to acknowledge the helpful advice of Professor Jan Schilt, Columbia University, in connexion with the necessity for taking account of the random motions.

XII. REFERENCES

BAADE, W., and MAYALL, N. U. (1951).—" Problems of Cosmical Aerodynamics." pp. 165–84. (U.S. Central Air Documents Office: Dayton, Ohio.)

BABCOCK, H. W. (1939).—*Bull. Lick Obs.* **19**: 41–51.

BOK, B. J. (1954).—Conference for instructors of astronomy, Leuschner Observatory, August-September 1954, lecture notes, p. 8.

BRACEWELL, R. N. (1955).—*Aust. J. Phys.* **8**: 54–60.

CHANDRASEKHAR, S. (1942).—" Principles of Stellar Dynamics." (Univ. Chicago Press.)

HOLMBERG, E. (1952).—Medd. Lunds Astr. Obs. Series 1, No. 180.

HOYLE, F. (1951).—" Problems of Cosmical Aerodynamics." pp. 195–7. (U.S. Central Air Documents Office: Dayton, Ohio.)

JEANS, J. H. (1922).—*Mon. Not. R. Astr. Soc.* **82**: 122–32.

KERR, F. J., HINDMAN, J. V., and ROBINSON, B. J. (1954).—*Aust. J. Phys.* **7**: 297–314.

KERR, F. J., and DE VAUCOULEURS, G. (1955).—*Aust. J. Phys.* **8**: 508–22.

KWEE, K. K., MULLER, C. A., and WESTERHOUT, G. (1954).—*Bull. Astr. Insts. Netherlds.* **12**: 211–22 (No. 458).

LOHMANN, W. (1954).—*Z. Astrophys.* **35**: 159–64.

MAYALL, N. U. (1948).—*Sky & Telesc.* **8**: 3–5, 17.

MAYALL, N. U. (1950).—*Publ. Obs. Univ. Mich.* **10**: 19–24.

MAYALL, N. U., and ALLER, L. H. (1942).—*Astrophys. J.* **95**: 5–23.

MAYALL, N. U., and EGGEN, O. J. (1953).—*Publ. Astr. Soc. Pacif.* **65**: 24–9.

OORT, J. H. (1940).—*Astrophys. J.* **91**: 273–306.

OORT, J. H. (1952).—*Astrophys. J.* **116**: 233–50.

PATTERSON, S. F. (1940).—Bull. Harv. Astr. Obs. No. 914, pp. 9–10.

PEREK, L. (1948).—*Contr. Astr. Inst. Masaryk Univ.* **1** (6): 1–34.

PEREK, L. (1950).—*Bull. Astr. Insts. Csl.* **2**: 75–9.

REDMAN, R. O. (1936).—*Mon. Not. R. Astr. Soc.* **96**: 588–604.

SCHWARZSCHILD, M. (1954).—*Astr. J.* **59**: 273–84.

SHAPLEY, H. (1950).—*Publ. Obs. Univ. Mich.* **10**: 79–84.

SHAPLEY, H., and NAIL, V. McK. (1955).—*Observatory* **75**: 120–2.

STRÖMGREN, B. (1952).—*Astr. J.* **57**: 65–83.

DE VAUCOULEURS, G. (1948).—*Ann. Astrophys.* **11**: 247–65.

DE VAUCOULEURS, G. (1953).—*Mon. Not. R. Astr. Soc.* **113**: 134–61.

DE VAUCOULEURS, G. (1954).—*Observatory* **74**: 23–31.

DE VAUCOULEURS, G. (1955*a*).—*Astr. J.* **60**: 126–40.

DE VAUCOULEURS, G. (1955*b*).—*Astr. J.* **60**: 219–30.

DE VAUCOULEURS, G. (1956).—*Occ. Notes R. Astr. Soc.* (in press).

VON WEIZSÄCKER, C. F. (1951).—*Astrophys. J.* **114**: 165–86.

WILSON, R. E. (1917).—*Publ. Lick Obs.* **13**: 187–90.

WYSE, A. B., and MAYALL, N. U. (1942).—*Astrophys. J.* **95**: 24–47.

ZWICKY, F. (1937).—*Astrophys. J.* **86**: 217–46.

LUNAR VARIATIONS IN THE IONOSPHERE

By R. A. Duncan*

[*Manuscript received October* 20, 1955]

Summary

The global pattern of the observed lunar variations of the height and electron density of the F_2 region is briefly summarized, new analyses being presented for Canberra (f_0F_2), Brisbane ($h'F_2$), and Washington (h' and f_0F_2). It is concluded that the height variation has an amplitude of from 1 to 3 km and a phase giving maximum height at 06 lunar hours at moderate latitudes and at 09 lunar hours at the geomagnetic equator. The critical frequency variation has an amplitude of from 2 to 4 per cent., maximum critical frequency occurring at about 09 lunar hours at moderate geomagnetic latitudes and 04 lunar hours at equatorial geomagnetic latitudes.

A theory of lunar ionospheric variations is then presented. The current system which Chapman has shown could be responsible for the observed lunar geomagnetic field variations is taken as a starting point: it is considered that this current must flow at a height of about 100 km. The tidal winds needed to drive the current, the potential distribution which will be set up in the dynamo layer, and the resulting periodic vertical drifts of ionization in the higher layers are calculated. It is shown that the divergence of drift velocity is too small to account for the lunar variations in f_0F_2. These are calculated taking into account the probable height variations of recombination coefficient and ionization production rate.

The conclusions thus reached are in good agreement with the observed variations in the F_2. It is concluded that the amplitude of the lunar tidal wind near the E layer is about 45 times greater than that observed on the ground.

I. Introduction

In recent years periodic atmospheric phenomena have been intensively studied. These phenomena include variations, in both solar and lunar time, of barometric pressure, of geomagnetic field, and of the maximum electron density in the ionosphere; and the horizontal and vertical movements of electron density peaks. The lunar variations can be safely ascribed to a single cause, the Moon's gravitational field, and in this paper an attempt will be made to relate ionospheric phenomena of lunar periodicity to this field.

In 1882 Balfour Stewart suggested that the daily magnetic variations were due to electric currents generated in the upper atmosphere by the daily convective movement of ionized air across the Earth's magnetic field. Subsequently Schuster (1908) and Chapman (1919) developed this " dynamo " theory quantitatively. Two unknowns were in the theory: the tidal velocity and the conductivity in the upper atmosphere.

---

* Radio Research Board, C.S.I.R.O., Electrical Engineering Department, University of Sydney.

From the study of barometric pressure oscillations, the lunar atmospheric tidal movements on the ground are fairly well understood, but there is no direct information about the corresponding movements in the upper atmosphere. The early workers, Laplace and Lamb, considering perfect isothermal and adiabatic atmospheres respectively, concluded that the tidal velocity at higher levels would be the same as at the ground, which is much too small for the purposes of the dynamo theory ; however, Pekeris (1937) showed that for other cases it was usual for the amplitude to increase with height. From a study of the vertical oscillation of the *E* layer, Martyn (1947) concluded that the amplification was probably about 200. A new estimate of the tidal velocity in the dynamo layer will be made in this paper.

Pedersen (1927) pointed out that the Earth's magnetic field impedes the movement of ions across it, and for this reason it has been difficult to find enough conductivity to account for the observed magnetic variations. However, Martyn (1948*b*) suggested that because of Hall current and consequent polarization the conductivity would not be as low as Pedersen had suggested, and later Baker and Martyn (1952, 1953) and, independently, Hirono (1950) and Fejer (1953) studied the problem of the conductivity of a thin layer of ionization in a magnetic field in detail and showed that, because of the form of the global wind pattern and the consequent inhibition of Hall current, vertical and horizontal polarization fields are set up, which increase the effective conductivity to about six times the Pedersen value over most of the Earth and to even higher values in a narrow strip over the geomagnetic equator. This removed the last objection to the dynamo theory.

The oscillations of height and ion density of the ionospheric layers remained to be explained. Martyn (1947) pointed out that these could hardly be attributed to the simple rising and falling of isobaric surfaces, and he suggested that they were due to electrodynamical interaction of the dynamo current with the geomagnetic field.

The revised dynamo theory of Baker and Martyn enables us to say with some certainty that the day-time dynamo layer is near the *E* region and to calculate the global distribution of electrical potential, which differs radically from that given by the Schuster theory. We shall here bring forward experimental evidence that notable lunar variations in ionospheric layer heights and electron densities occur chiefly in the day-time, when the *E* layer is present. These facts permit the presentation of a consistent explanation of lunar variations.

We first describe the observed ionospheric lunar variations. Then, taking the current system which Chapman has shown to be consistent with the observed lunar geomagnetic field variations, we calculate the necessary tidal winds, the potential field set up in the dynamo layer, and the periodic movements of ionization and changes in ion density which this field will produce in higher layers. The conclusions thus reached are reconciled with all available measurements of movements and electron density changes in the F_2 region of the ionosphere.

II. THE OBSERVED LUNAR VARIATIONS OF ELECTRON DENSITY AND HEIGHT OF THE F_2 REGION

Martyn (1947, 1948*a*), Appleton and Beynon (1948), Burkard (1948), and Matsushita (1949) have shown that the main lunar periodicity of F_2 electron density and height is semi-diurnal. Maximum heights generally occur at 06 lunar hours at moderate latitudes and at 08 lunar hours above the geomagnetic equator (Table 1), the only serious exceptions to this rule being Ottawa and Watheroo, where the tidal amplitudes are small and the phase determinations therefore not so reliable. Maximum electron densities occur 3–4 hr after maximum height at moderate latitudes, and about 4 hr before maximum height at Huancayo.

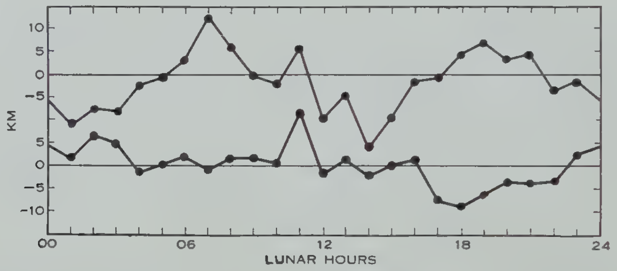


Fig. 1.—The lunar variation of $h^{\text{max.}}F_2$ at 13 (top) and 01 (bottom) solar hours. Huancayo 1942–43–44.

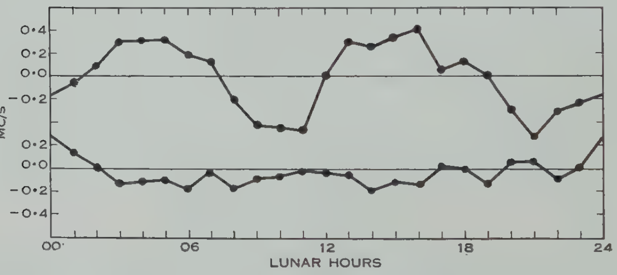


Fig. 2.—The lunar variation of f_0F_2 at 11 (top) and 23 (bottom) solar hours. Huancayo 1942–43–44.

Analyses made by the Radio Research Board, which will be published in detail later, show that large lunar semi-diurnal variations of $h^{\text{max.}}$ and f_0F_2 occur above Huancayo only during the day (Figs. 1 and 2), and above Canberra only during the late afternoon and evening and for a very brief period at dawn.

III. THE THEORY OF LUNAR IONOSPHERIC VARIATIONS

(*a*) *The Lunar Dynamo Current*

Chapman and Bartels (1940) have calculated the form and magnitude of the currents which must flow in the ionosphere if the observed lunar geomagnetic variations are to be explained in this manner. Figure 3 shows the current at new moon during the equinox: a similar current system centred near solar noon

TABLE 1

HARMONIC COEFFICIENTS, P_2 AND t_2, OF THE SEMI-DIURNAL LUNAR VARIATIONS OF F_2 REGION DATA

n is the number of months' data examined. Martyn's computations have been re-checked using slightly more accurate procedures, and differences, mostly trivial, will be found between the old values and those given here. Martyn's analyses for f_0F_2 Canberra, $h'F_2$ Brisbane, and f_0 and $h'F_2$ Washington have been extended to cover twice the previous number of months' data

Station	Geomagnetic Latitude	Investigator	$h'F_2$			$h^{\text{max.}}F_2$			f_0F_2		
			P_2 (km)	t_2	n	P_2 (km)	t_2	n	P_2 (Mc/s)	t_2	n
Huancayo	− 0·6	Martyn	2·3	9·2	36	5·2	8·4	30	0·14	4·3	48
Cape York	−20·7	Martyn	2·0	6·7	11	2·2	6·3	6	0·12	9·7	10
Kihei	+20·9	Martyn	2·6	7·9	15				0·17	11·4	14
Tokyo	+25·5	Matsushita	3·0	6·6	13						
Durbanville	−32·6	Martyn							0·05	11·6	8
Brisbane	−35·7	Martyn	1·3	6·4	82	2·0	6·0	33	0·08	9·3	31
Watheroo	−41·7	Martyn	0·3	5·9	85	0·6	8·0	73	0·03	9·9	112
Canberra	−44·0	Martyn	1·6	5·6	48	1·7	5·6	36	0·06	9·0	72
Christchurch	−48·0	Martyn	0·8	5·2	28				0·05	8·5	28
Washington	+50·3	Martyn	1·4	5·8	77				0·04	10·2	77
Washington	+50·3	Burkard							0·09	11·0	12
Hobart	−51·6	Martyn	2·0	4·6	10	1·1	5·6	9	0·04	9·3	10
Slough	+54·3	Appleton & Beynon				2·0	6·0		0·05	11·0	12
Ottawa	+56·9	Martyn	0·6	3·0	24				0·01	11·2	30
Burghead	+59·4	Martyn							0·02	8·3	48

in each case flows at first quarter, full moon, and last quarter. At intermediate lunar ages, there is slight distortion of this picture, the day-night boundary cutting across the current loops. As a result of more recent geomagnetic analyses, it is known that an anomalously large current of about twice the intensity shown in Figure 3 flows above the geomagnetic equator.

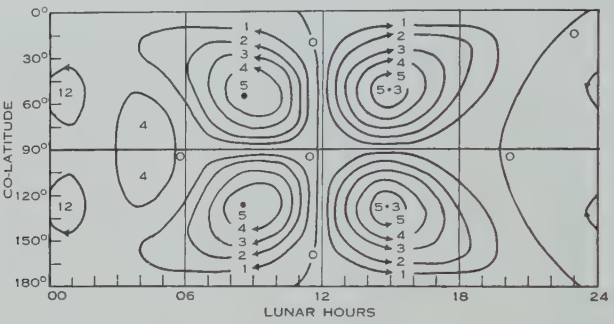


Fig. 3.—The system of ionospheric currents which could cause the lunar geomagnetic variations observed at new moon during the equinoxes. Contours 1000 A apart (after Bartels).

If we ignore the difference between the day and night intensities and the equatorial anomaly, the current system of Figure 3 is well described by the function :

$$R = 2\cdot 60 J \sin^2\theta \cos\theta \sin 2\varphi \text{ e.m.u.}, \quad \ldots\ldots\ldots\ldots \quad (\mathbf{1})$$

$$j_\theta = \frac{\partial R}{r \sin\theta \partial\varphi}, \quad j_\varphi = -\frac{\partial R}{r\partial\theta},$$

where θ is the co-latitude

φ is the longitude measured from the Moon,

r is the radius of the Earth,

j_θ is the southward current density,

j_φ is the eastward current density.

This function represents a current system with vortex centres at $\varphi = 45° \pm m90°$ and $\theta = 54\cdot 7°$ and $125\cdot 3°$, and with a circulation per vortex of J. A clockwise vortex with a day-time circulation of 500 e.m.u. flows in the first octant (Fig. 3) so that J has a day-time value of -500. We shall take the equatorial eastward current j_e to be double that given by (**1**), so that

$$j_e = (5\cdot 20/r) J \sin 2\varphi. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{2})$$

The E region must carry the substantial part of this current during the day and early evening, and, as detectable lunar variations in the F_2 are almost entirely confined to these hours, we shall not concern ourselves with the problem of the height distribution of the relatively small night-time dynamo currents.

(b) *Lunar Tidal Winds in the Dynamo Layer*

Barometric observations show a lunar pressure variation (p) on the ground of the form

$$p = \sin^3 \theta \cos 2\varphi. \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Fejer (1953) was the first to point out that the Earth's rotation enhances the electromotive effect of the tidal winds, as Coriolis deflection induces east-west air motion, i.e. air motion normal to the geomagnetic field, at higher latitudes. Taking account of the Earth's rotation, then, southward and eastward wind components

$$\left.\begin{aligned} v_\theta &= -\cos\theta \sin 2\varphi, \\ v_\varphi &= -(1-0\cdot59 \sin^2\theta)\cos 2\varphi \end{aligned}\right\} \qquad \ldots\ldots\ldots (4)$$

may be expected (Gold 1910). We are free to choose a suitable phase and amplitude for these winds in the dynamo layer.

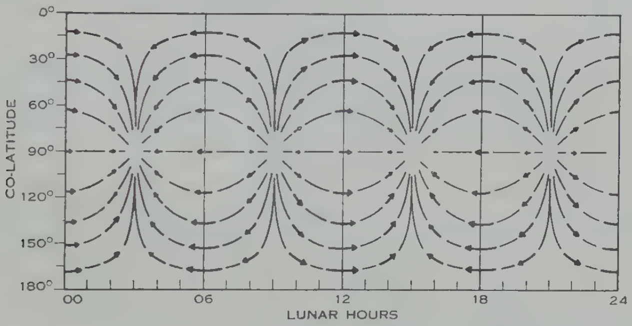


Fig. 4.—The global distribution of the lunar gravitational tidal winds in the dynamo region. An arrow one division long represents a velocity of 100 cm/sec.

Although the conduction of the ionosphere depends on the formation of Hall polarization, Baker and Martyn have shown that the dynamo current at high latitudes is parallel to the dynamo electromotive force, and maximum eastward wind must therefore occur at the time of maximum poleward current, that is at 00 and 12 lunar hours. The phase of the winds in the E layer must therefore be opposite to that on the ground with components of the form

$$\left.\begin{aligned} v_\theta &= v_{\text{max.}} \cos\theta \sin 2\varphi, \\ v_\varphi &= v_{\text{max.}} (1-0\cdot59 \sin^2\theta)\cos 2\varphi. \end{aligned}\right\} \qquad \ldots\ldots\ldots (5)$$

The wind velocities may be estimated by equating the dynamo electromotive force to the ohmic potential drop around a circuit which runs along the meridian $\varphi=0$ from the equator to the pole, back to the equator along the meridian $\varphi=\frac{1}{2}\pi$, and then along the equator to the starting point: this corresponds very closely to the circuit of maximum dynamo voltages and currents. The dynamo field E_θ along the meridians is

$$E_\theta = H_r v_\varphi, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

where H_r, the vertical component of the geomagnetic field, equals $-0\cdot622\cos\theta$. No dynamo field is generated along the equator and the total e.m.f. around the circuit is therefore

$$\text{e.m.f.}=2\int_0^{\frac{1}{2}\pi}-0\cdot622\cos\theta\, v_{\max.}(1-0\cdot59\sin^2\theta)r d\theta$$

$$=1\cdot00 r v_{\max.}\qquad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\quad(7)$$

We shall accept Martyn and Baker's value of $3\cdot5\times10^{-8}$ e.m.u. for the conductivity in the non-equatorial region so that the ohmic drop along the meridians is

$$\frac{2}{3\cdot5\times10^{-8}}\int_0^{\frac{1}{2}\pi} j_\theta r d\theta,$$

i.e. from (**1**), equal to $(1\cdot49\times10^8)J$. Along the equator the conductivity is $1\cdot64\times10^{-7}$, so that from (**2**) the potential drop is

$$\frac{5\cdot20J}{r(1\cdot64\times10^{-7})}\int_{\frac{1}{2}\pi}^{0} r\sin 2\varphi d\varphi=(1\cdot58\times10^7)J.$$

Hence the total potential drop around the circuit is

$$1\cdot65\times10^8 J.\qquad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\quad(8)$$

Equating (**7**) and (**8**), we find

$$v_{\max.}=130\text{ cm/sec.}\qquad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\quad(9)$$

This is about 45 times as great as the lunar tidal wind observed on the ground. The E region winds are shown in Figure 4.

(c) *Polarization and Electric Fields in the Ionosphere*

The E-layer dynamo current will induce polarization, the horizontal distribution of which can be most easily described by dividing it into two parts, an equatorial Schuster polarization and a Hall polarization at the centre of each current loop.

The intensity of this polarization is independent of the diurnal variations of conductivity, as the dynamo current is proportional to the ion density and the induced polarization is proportional to this current divided by the ion density. We may therefore use noon values of current and conductivity in our computations without loss of generality.

The latitude variation of conductivity arising from the variation of solar zenith angle and magnetic dip is important. Although a more precise investigation is by no means intractable, we shall adopt the simplified picture of Baker and Martyn (1953): a region lying within 7° of the geomagnetic equator is taken to have a height integrated direct east-west noon conductivity $\Sigma_y=1\cdot64\times10^{-7}$ e.m.u. Above the rest of the Earth, a direct conductivity $\Sigma_1=6\cdot4\times10^{-9}$ e.m.u., a Hall conductivity $\Sigma_2=+1\cdot36\times10^{-8}$ e.m.u. in the northern and $-1\cdot36\times10^{-8}$ e.m.u. in the southern hemisphere, and an effective conductivity $\Sigma_3 = \Sigma_1 + \Sigma_2^2/\Sigma_1 = 3 \cdot 5 \times 10^{-8}$ e.m.u. will be assumed. The meaning of these four conductivities has already been described by Baker and Martyn and will be clear from the context.

The dynamo electromotive forces (Fig. 5) do not form closed circuits but converge on, or diverge from, four points spaced around the equator so causing " Schuster " polarization, the geographical distribution of which may be described approximately by the expression

$$V_s = V_s^{\text{max.}} \sin^2 \theta \cos 2\varphi. \quad \ldots\ldots\ldots\ldots \quad (10)$$

It is this polarization which closes the current loops by producing flow along the equatorial parallels of latitude so that

$$j_e = E_\varphi \Sigma_y, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (11)$$

where E_φ, the eastward field, equals

$$-\frac{\partial V_s}{r \sin \theta \partial \varphi} = 2 V_s^{\text{max.}} \frac{\sin \theta}{r} \sin 2\varphi.$$

Therefore, from (2)

$$\frac{5 \cdot 20 J}{r} = \frac{2 V_s^{\text{max.}} \Sigma_y}{r}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

i.e. $V_s^{\text{max.}} = -79 \cdot 3 \times 10^8$ e.m.u. and the extreme potentials due to horizontal polarization above the geomagnetic equator are ±80 V.

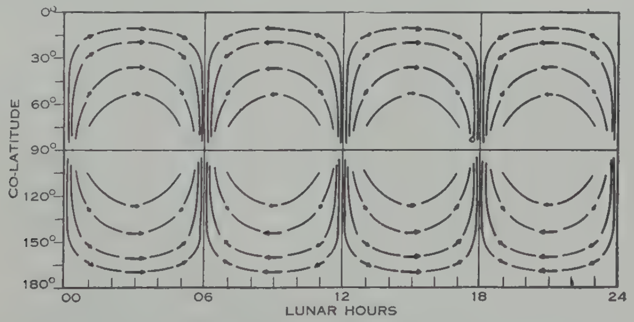


Fig. 5.—The global distribution of the electromotive forces in the dynamo region. An arrow one division long represents an electromotive field of 100 mV/km.

Because the ionosphere has a Hall conductivity, initially a component of current will flow at right angles to the dynamo electromotive forces, but as such currents all converge (or diverge) from the vortex centres, they will be rapidly countered by the accumulation of Hall polarization. As is explained in Baker and Martyn's papers, this Hall polarization then causes its own Hall current to flow around the current loop, augmenting the dynamo current and raising the effective conductivity from Σ_1 to $\Sigma_1 + \Sigma_2^2/\Sigma_1$. Of the observed non-equatorial current, therefore, a part proportional to Σ_1 is due directly to the dynamo voltages, but the major part, proportional to Σ_2^2/Σ_1 (and thus a fraction $\Sigma_2^2/(\Sigma_1^2+\Sigma_2^2)=0\cdot 83$ of the whole) is caused indirectly by the Hall polarization. The Hall current function R_H is therefore from (**1**)

$$\left.\begin{aligned} R_H &= 0\cdot 83R, \\ j_{H,\theta} &= 0\cdot 83\frac{\partial R}{r\sin\theta\partial\varphi}, \\ j_{H,\varphi} &= -0\cdot 83\frac{\partial R}{r\partial\theta}. \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{13})$$

This current is related to the Hall field E_H and polarization potential V_H by the relations

$$\left.\begin{aligned} j_{H,\theta} &= \Sigma_2 E_{H,\varphi} &&= -\Sigma_2\frac{\partial V_H}{r\sin\theta\partial\varphi}, \\ j_{H,\varphi} &= -\Sigma_2 E_{H,\theta} &&= +\Sigma_2\frac{\partial V_H}{r\partial\theta}. \end{aligned}\right\} \qquad \ldots\ldots \quad (\mathbf{14})$$

Hence V_H varies in the same manner as R, that is,

$$V_H = 2\cdot 60\,V_H^{\text{max.}}\sin^2\theta\cos\theta\sin 2\varphi, \qquad \ldots\ldots \quad (\mathbf{15})$$

$$V_H^{\text{max.}} = \frac{-0\cdot 83J}{\Sigma_2} = \pm 3\cdot 05\times 10^{10} \qquad \ldots\ldots\ldots\ldots \quad (\mathbf{16})$$

in the northern and southern hemispheres respectively.

The complete expression for the global distribution of polarization potential V in the dynamo layer is therefore

$$V = [\pm 300\times 2\cdot 60\sin^2\theta\cos\theta\sin 2\varphi - 80\sin^2\theta\cos 2\varphi]\times 10^8 \text{ e.m.u.} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{17})$$

This is shown in Figure 6.

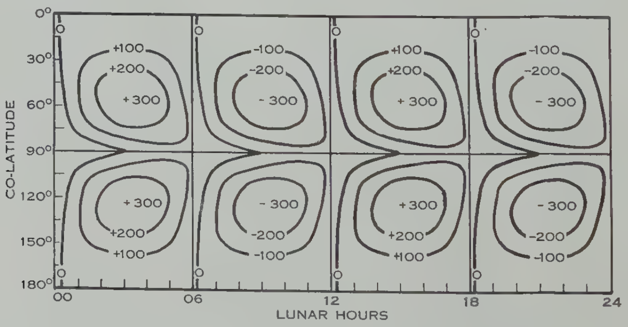


Fig. 6.—The global distribution of horizontal polarization in the dynamo region. Contours 100 V apart.

The polarization of the E region will be communicated to the higher levels along the highly conducting and thus equipotential geomagnetic field lines. Hence, generally,

$$V = [\pm 780\sin^2\theta_E\cos\theta_E\sin 2\varphi - 80\sin^2\theta_E\cos 2\varphi]\times 10^8, \qquad \ldots\ldots \quad (\mathbf{18})$$

where θ_E is the co-latitude at which the region connects via the geomagnetic field lines with the dynamo layer. From the properties of the geomagnetic field

$$\sin^2 \theta_E = (r_E/r) \sin^2 \theta, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(19)}$$

so that the potential function for the complete ionosphere is

$$V = [\pm 780 (r_e/r) \sin^2 \theta \sqrt{\{1-(r_E/r)\sin^2\theta\}} \sin 2\varphi - 80 (r_E/r) \sin^2 \theta \cos 2\varphi] \times 10^8,$$
$$V \simeq [\pm 780 \sin^2 \theta \sqrt{\{1-(r_E/r)\sin^2\theta\}} \sin 2\varphi - 80 \sin^2 \theta \cos 2\varphi] \times 10^8. \quad \ldots \mathbf{(20)}$$

In the E region the total electric field will be the sum of the electromotive and electrostatic fields, but due to the greatly decreased pressure at higher levels, which will entail a much greater reduction of gravitational tidal power per unit volume than of viscous and electrodynamic damping, it is likely that in the F region gravitational tidal winds are negligible and that the electrostatic field is the only one which need be considered. There will be horizontal air movements in the F_2 due to the interaction of the electric currents produced by the electrostatic field and the geomagnetic field, but, as Baker and Martyn (1953) have shown, the e.m.f.'s developed by these will be back e.m.f.'s and will merely reduce the effective electric field. This effect will be discussed later.

(*d*) *Vertical Drift of Ionization in the* F$_2$ *Region*

Martyn has shown that, because of the geomagnetic field, an electric field in the ionosphere produces a drift of neutral ionization. This drift is the same as that given by Ampere's law; an eastward current being associated with an upward drift of ionization.

A discussion of ion drift in the E region is difficult because of the very rapid change with height, in this region, of the conductivity per ion pair; we shall confine ourselves to the very much simpler problem of ion drift in the F_2 region. Here there is no Hall current so that an eastward field will cause an eastward current. Martyn (1953) has derived the relation

$$v_r = \frac{E\varphi}{H} \cos \psi \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(21)}$$

for the upward drift velocity (v_r), where $E\varphi$ is the eastward electric field, H the geomagnetic field intensity, and ψ the geomagnetic dip angle.

As has been pointed out (Baker and Martyn 1953), this relation will need modification if the air is set in motion by the drifting ionization. A rough measure of the time which air will take to acquire the velocity of ionization drifting through it is given by $t = \rho/(\sigma H^2)$, where ρ is the air density and σ the conductivity. If we evaluate this for the F_2 region we find that t is of the order of an hour; independent of the air pressure. Hence some air motion probably occurs. However, vertical air drift will be prevented by the Earth's gravitational field and even horizontal motion will be affected by a redistribution of air pressure because the horizontal component of the ion drift has appreciable divergence. Air motion is therefore probably a second order effect, but it should be remembered that it will reduce the vertical ion drift to some extent, particularly at high latitudes.

Going back to equation (**21**), we have from the properties of the geomagnetic field

$$\frac{\cos\psi}{H}=\frac{\sin\theta}{0\cdot 311(1+3\cos^2\theta)},$$

also,

$$E\varphi=-\frac{\partial V}{r\sin\theta\partial\varphi},$$

so that

$$v_r=-\frac{\partial V}{0\cdot 311r(1+3\cos^2\theta)\partial\varphi}. \qquad \dots\dots\dots\dots (\mathbf{22})$$

From (**20**), then

$$v_r=80\cdot 3\frac{\sin^2\theta\sqrt{(94\cdot 5\cos^2\theta_E+1)}}{1+3\cos^2\theta}\left[\cos\left(2\varphi-\arctan\frac{-1}{-9\cdot 75\cos\theta_E}\right)\right]. \qquad \dots\dots\dots\dots (\mathbf{23})$$

Hence maximum upward drift velocity should occur at 06 lunar hours over most of the Earth. The theoretical variation of the amplitude of the drift with latitude is shown in Figure 7. Above the geomagnetic equator the time of greatest upward drift depends on the height, being 09 hours in the dynamo region (100 km), 07 hours at 300 km, and approaching 06 lunar hours at very great heights.

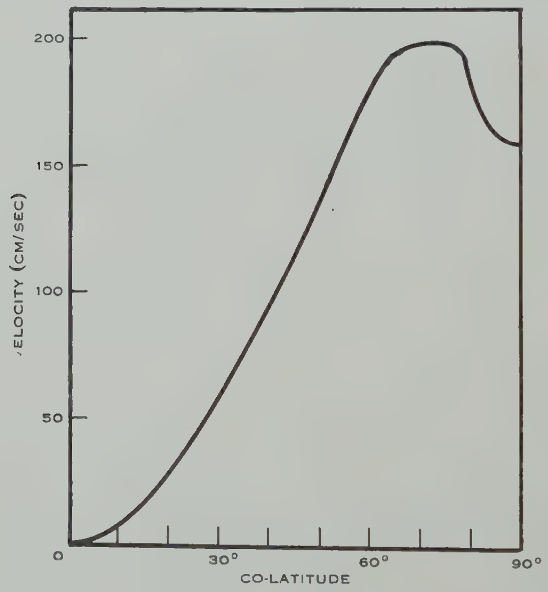


Fig. 7.—The variation with co-latitude of the vertical lunar drift velocity at 300 km.

(*e*) *Lunar Periodicity in the Height of the* F *Region*

Integration of equation (**23**) shows that if the *F* region were translated up and down by the drift it would oscillate through about 12 km, maximum height being reached 3 hr after the time of greatest upward velocity. However, this large movement does not occur because the limited life of electrons (about an hour in the day-time F_2 region) prevents them from being shifted very far from the height at which they are produced. Even at night diffusion and the height gradient of the recombination coefficient tend to stabilize the height of the F_2 layer : if the layer is shifted upwards it tends to diffuse down again : if the layer is shifted downwards the accelerated decay of the lower edge will tend to restore it to its original height. Vertical drift, therefore, will perturb the height of an ionospheric layer rather than shift it bodily.

Suppose that we have in an ionospheric region, ionization, recombination, and a redistribution of electrons by a solar tide. The equation of continuity of the electron density is

$$\frac{\partial N}{\partial t}=I-\alpha N^2-v_r\frac{\partial N}{\partial h}-N\frac{\partial v_r}{\partial h}, \qquad \ldots\ldots\ldots\ldots\ldots \text{(24)}$$

where N is the electron density, t is time, I is the rate of ionization, α is the recombination coefficient, v_r is the velocity of vertical drift, and h is height. In time the electron density will approach a value N_e such that $\partial N_e/\partial t=0$. If the electron density is perturbed from this value by an amount n so that $N=N_e+n$, and the perturbing cause then removed, the equation of continuity becomes

$$\frac{\partial n}{\partial t}=-2\alpha N_e n-n\frac{\partial v_r}{\partial h}-v_r\frac{\partial n}{\partial h}.$$

The only term on the right of this equation which represents a tendency for the perturbation to decay is the first, $-2\alpha N_e n$, the other two terms merely show that the electron density perturbation is redistributed by the solar tide. The electron density perturbation n decays, then, according to the law

$$\partial n/\partial t=-2\alpha N_e n,$$

or

$$n=\exp\left(-2\alpha N_e t\right), \qquad \ldots\ldots\ldots\ldots\ldots \text{(25)}$$

so that $\tau=1/2\alpha N$ may be taken as the " relaxation time " of the ionosphere.

If this relaxation time is much shorter than 3 hr, we should expect an ionospheric layer to be displaced a distance $\Delta h=v_r\tau$ by a vertical drift v_r. Of all time intervals τ, that beginning $\frac{1}{2}\tau$ seconds before and ending $\frac{1}{2}\tau$ seconds after the time of maximum upward drift, covers the period of greatest upward drift, and the electrons should be at their greatest height $\frac{1}{2}\tau$ seconds after the time of maximum upward velocity. The following more rigorous treatment shows this picture to be correct.

If $\mathrm{d}N/\mathrm{d}h$ is the height gradient of the electron density, then in the presence of a uniform vertical drift v_r the equation of continuity becomes

$$\frac{\partial N}{\partial t}=I-\alpha N^2-v_r\frac{\partial N}{\partial h},$$

that is,

$$\frac{\partial N}{\partial t}=-2\alpha Nn-v_r\frac{\partial N}{\partial h}, \qquad \ldots\ldots\ldots\ldots\ldots \text{(26)}$$

When the relaxation time is short, the layer will reach a new equilibrium state so that $\partial N/\partial t=0$,

and hence

$$n=v_r\tau\frac{\partial N}{\partial h}. \qquad (27)$$

This relation implies that to the first order we can consider the ion density perturbation at a given height to be due to the layer being raised undistorted a distance Δh where

$$\Delta h=v_r\tau. \qquad (28)$$

(It should perhaps be pointed out that the electron density perturbation at a given height is deduced in (**27**) merely as a means of obtaining the height perturbation. It is not the observed quantity, the perturbation of the value of the maximum electron density. The latter will be calculated in the next section.)

It is difficult to obtain reliable estimates of τ in the F_2 but if we accept S. K. Mitra's (1952) values of recombination coefficient, we obtain a midday figure of 4000 sec. From this we should expect a day-time lunar height variation of about 6 km, maximum height occurring at 06 lunar hours at moderate latitudes and 07 lunar hours at the geomagnetic equator. Even larger amplitudes and later phases would be expected at night. However, we shall show that at F_2 heights, and particularly at the height of the night F_2 (300 km) diffusion becomes a process important enough to keep the effective relaxation time of the ionosphere short.

The greatest height at which reasonably reliable rocket measurements of temperature and mean free paths have been made is 200 km (Rocket Panel 1952), and on this evidence the coefficient of diffusion of air (K) at this level is roughly $7\cdot5\times10^8$ cm²/sec. The diffusion coefficient will increase with height due to the increasing temperature and decreasing pressure. Pressure decreases exponentially with height and will have much the greater influence so we shall write

$$K=(7\cdot5\times10^8)e^{h/H}, \qquad (29)$$

where h/H is the reduced height above our 200 km datum level.

Huxley (1952) and others before him have shown that the electrons and positive ions in the ionosphere are held together by electrostatic attraction, the electron-positive ion gas diffusing at twice the rate appropriate to the positive ions alone. For our purposes, it will be sufficiently accurate to assume the coefficient of diffusion of the latter equal to that of neutral molecules, so that the coefficient of diffusion of the electron-positive ion gas (K_i) is

$$K_i=(1\cdot5\times10^9)e^{h/H}. \qquad (30)$$

The equation of continuity for the electron-positive ion gas is

$$\frac{\partial N}{\partial t}=\mathrm{div}\left[K_i\frac{\partial N}{\partial h}-WN\right], \qquad (31)$$

where W is the drift velocity of the ions under gravity. It may be shown from kinetic theory that

$$W=-\frac{K_i mg}{kT}=-\frac{K_i}{H_i}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (32)$$

so that, as is well known, diffusion and gravitational drift are in equilibrium when

$$N=N_0 \exp\left(-h/H_i\right), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (33)$$

where H_i, the scale height for the electron-positive ion gas, will be twice that for the neutral molecules (H).

It is clear, however, that in an ionospheric layer gravitational drift and diffusion will not generally balance one another ; indeed, on the under side of the layer, the side observed, they both cause a downward flux of ionization. The total effective velocity of this flux (U) is

$$U=-K_i\frac{\partial N}{N\partial h}+W,$$

or

$$U=-K_i\left(\frac{\partial N}{N\partial h}+\frac{1}{H_i}\right). \quad \ldots\ldots\ldots\ldots\ldots \quad (34)$$

We have not yet considered the effect of the geomagnetic field. Ions can gravitate and diffuse only in the direction of this field. If the dip is ψ the components of ion density gradient and gravitational field in this direction are the corresponding vertical components multiplied by $\sin\psi$. To find the vertical component of velocity we must multiply by $\sin\psi$ once again, so that

$$\left.\begin{aligned} U&=-K_i\left(\frac{\partial N}{N\partial H}+\frac{1}{H_i}\right)\sin^2\psi, \\ \text{that is,}\quad U&=-2K\left(\frac{\partial N}{N\partial h}+\frac{1}{2H}\right)\sin^2\psi. \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (35)$$

At night we may consider the F_2 layer to consist of a relatively thick bank of ionization in the upper region of rapid diffusion and low decay, with a lower boundary whose position is determined by the opposing processes of downward diffusion and gravitational drift on one hand, and the relatively rapid decay of the lower edge on the other. A reasonable value for the scale height of the F_2 region is 30 km. We shall take the semi-thickness of the F_2 to be 60 km so that the mean value of $\partial N/\partial h$ is $N_{\text{max.}}/(6\times10^6)$. The mean value of N may be taken as $\frac{1}{2}N_{\text{max.}}$, so that

$$U=-(1\cdot5\times10^9)\left[\frac{1}{3\times10^6}+\frac{1}{6\times10^6}\right]\sin^2\psi e^{h/H},$$

that is,

$$U=-750\sin^2\psi e^{h/H}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (36)$$

and hence

$$\frac{dU}{dh}=-25\times10^{-5}\sin^2\psi e^{h/H}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (37)$$

An upward electrodynamic drift (v_r) will push the F_2 layer up until the downward diffusion velocity ($-U$) is increased by an amount equal to v_r. Hence, if in the relation

$$dh=-\frac{10^5}{25}\operatorname{cosec}^2\psi e^{-h/H}dU \quad \ldots\ldots\ldots\ldots\ldots \quad (38)$$

we put $-dU$ equal to the upward drift v_r, dh is the displacement which this drift will cause, that is,

$$\Delta h=\frac{10^5}{25}\operatorname{cosec}^2\psi e^{-h/H}v_r, \quad \ldots\ldots\ldots\ldots\ldots \quad (39)$$

or

$$\Delta h=\frac{10^5}{25}\,\frac{1+3\cos^2\theta}{4\cos^2\theta}e^{-h/H}v_r. \quad \ldots\ldots\ldots\ldots\ldots \quad (40)$$

The F_2 layer is about two scale heights above the 200 km level. We shall put therefore,

$$h/H=2,$$

and

$$v_r=155 \text{ cm/sec},$$

corresponding to the situation at co-latitude 55°.

This gives $\Delta h=1\cdot 26$ km.

This crude argument does not take account of the height gradient of recombination, but it indicates approximately how much diffusion will limit the amplitude of the lunar height variations.

It seems therefore that, at moderate latitudes, the amplitude of both the day and night lunar variation of $h'F_2$ should be about a kilometre, and that maximum h' and $h^{\text{max.}}F_2$ should occur close to the time of maximum upward drift velocity ; 06 lunar hours. Table 1 shows that, at most moderate latitude stations, maximum F_2 heights occur close to 06 lunar hours, and that the amplitude of the variation is from 1 to 3 km, in fair agreement with these arguments.

Above the geomagnetic equator vertical diffusion is inhibited and this will lead to large F_2 height variations. The maximum drift velocity here is 160 cm/sec, so that at midday when the relaxation time is about 4000 sec, a maximum positive height variation of about 6 km should occur close to the time of maximum upward drift ; 07 lunar hours. Towards sunset, when the relaxation time is of the order of hours, even larger amplitudes and a phase of 10 lunar hours can be expected.

Martyn (1947) has shown that the observed variation at Huancayo does in fact exhibit these features.

(*f*) *Lunar Periodicity in the Electron Density of the* F$_2$ *Region*

Periodic drift may result in periodic variations in electron density in two ways :

(i) The drift velocity may have a gradient, resulting in the periodic concentration and rarefaction of the ionosphere.

(ii) The drift may shift the layer to a height where a different ionization and recombination rate prevails.

Let us first consider the effect of drift velocity gradients. The drift velocity **v** is given in magnitude and direction by the relation :

$$\mathbf{v}=\frac{1}{H^2}(\mathbf{E}\times\mathbf{H}). \qquad \ldots\ldots\ldots\ldots \quad (41)$$

The divergence of the drift is therefore,

$$\begin{aligned}\text{div }\mathbf{v}&=\nabla\cdot\left[\frac{1}{H^2}(\mathbf{E}\times\mathbf{H})\right]\\ &=\frac{1}{H^2}\nabla\cdot(\mathbf{E}\times\mathbf{H})+\nabla\frac{1}{H^2}\cdot(\mathbf{E}\times\mathbf{H})\\ &=\frac{1}{H^2}\,(\text{curl }\mathbf{E}.\mathbf{H}-\mathbf{E}.\text{curl }\mathbf{H})+\left(\text{grad }\frac{1}{H^2}\right).(\mathbf{E}\times\mathbf{H}).\end{aligned}$$

Now the first term is zero as both **E** and **H** can be derived from scalar potentials so that

$$\text{div }\mathbf{v}=\left(\text{grad }\frac{1}{H^2}\right).(\mathbf{E}\times\mathbf{H}). \qquad \ldots\ldots\ldots\ldots \quad (42)$$

As both H_φ and $\partial(1/H^2)/\partial\varphi$ are zero, this reduces to

$$\text{div }\mathbf{v}=E_\varphi\left[-\frac{\partial(1/H^2)}{\partial r}H_\theta+\frac{\partial(1/H^2)}{r\partial\theta}H_r\right]. \qquad \ldots\ldots\ldots\ldots \quad (43)$$

The geomagnetic field is such that

$$H_r=\frac{2M\cos\theta}{r^3},$$

$$H_\theta=\frac{M\sin\theta}{r^3},$$

$$H^2=\frac{M^2}{r^6}\ (3\cos^2\theta+1).$$

where M is the Earth's magnetic moment. Hence,

$$\frac{\partial(1/H^2)}{\partial r}=\frac{6r^5}{M^2\,(3\cos^2\theta+1)},$$

and

$$\frac{\partial(1/H^2)}{r\partial\theta}=\frac{6r^5\cos\theta\sin\theta}{M^2\ (3\ \cos^2\ \theta+1)^2}.$$

Substituting these values in (43) we find that

$$\text{div }\mathbf{v}=-\frac{6r^2\sin\theta\,(\cos^2\theta+1)}{M\,(3\cos^2\theta+1)^2}E_\varphi. \qquad \ldots\ldots\ldots\ldots \quad (44)$$

It is useful to obtain this expression in terms of the vertical component of drift (v_r). From (21)

$$v_r=-\frac{r^3\sin\theta}{M\,(3\cos^2\theta+1)}E_\varphi,$$

therefore

$$\operatorname{div} \mathbf{v} = +\frac{6\,(\cos^2\theta + 1)}{r\,(3\cos^2\theta + 1)} v_r.$$

$(1/N)(\partial N/\partial t) = -\operatorname{div} \mathbf{v}$ and, if the F_2 relaxation time is τ, the critical frequency variation will be

$$\frac{\Delta f_0}{f_0} = -\frac{3(1+\cos^2\theta)}{r(1+3\cos^2\theta)} v_r \tau = -\frac{3(1+\cos^2\theta)}{r(1+3\cos^2\theta)} \Delta h. \quad \ldots . \ (45)$$

Thus the critical frequency variation due to drift divergence will be in phase opposition to the height variation.

Martyn (1955) has calculated the vertical gradient of the vertical component of the drift velocity $(\partial v_r/\partial r)$ on the assumption that the geomagnetic field intensity is uniform, and he has equated this to $-\partial N/N\partial t$. Clearly, we should also consider the gradient of the eastward component of drift $(\partial v_\varphi/r \sin\theta\partial\varphi)$. The new term is equal and opposite to the old (i.e. $\partial v_r/\partial r = -\partial v_\varphi/r\sin\theta\partial\varphi$) so that the total velocity divergence is zero. The derivation given here [(**41**)–(**45**)] shows that this is generally true: the drift divergence due to the electric field gradient is zero because $\operatorname{div}(\mathbf{E}\times\mathbf{H}) \equiv 0$. The residual velocity divergence is due to the gradient of the geomagnetic field intensity.

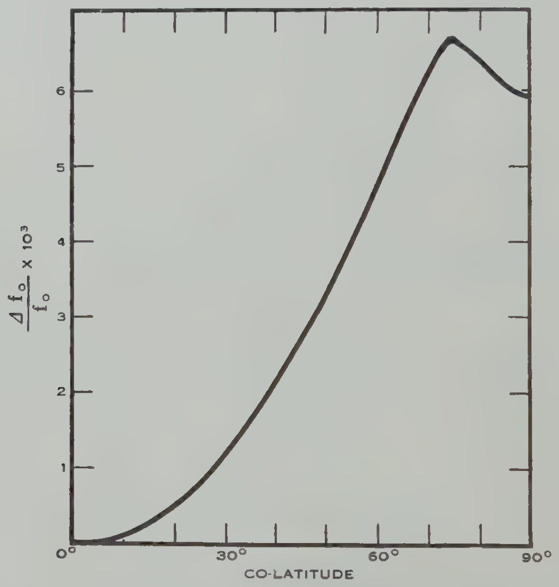


Fig. 8.—The variation with co-latitude of the relative critical frequency changes due to lunar drift velocity divergence at 300 km.

The critical frequency variation given by equation (**45**), for a height of 300 km, τ equal to 4000 sec, and v_r as given in (**23**), is plotted against geomagnetic latitude in Figure 8. It is seen that a critical frequency variation of about $\frac{1}{2}$ per cent. may result near the geomagnetic equator, but that generally the effect is small.

It seems likely, therefore, that electron density variations at moderate latitudes are principally due to the effect of drift in an ionosphere with a height gradient of electron production and recombination.

Let I be the electron production rate,
α be the recombination coefficient,
N_m be the maximum electron density,
h_m be the height at which this occurs,
$N_{m,0}$ be the maximum electron density which would prevail if there were no drift, and
$h_{m,0}$ be the height at which this would occur, and
$N_{e,0}$ be the equilibrium electron density (i.e. $N_{e,0}=\sqrt{(I_0/\alpha)}$) at the height $h_{m,0}$.

Then let

$$n=N_m-N_{e,0},$$
$$h'=h_m-h_{m,0}.$$

h' is the tidal perturbation of h_m, but n, the perturbation of the maximum electron density, will be due partly to the drift and partly to the natural difference between the actual electron density and the equilibrium value at any time, that is,

$$n=N_m-N_{e,0}=(n_m-N_{m,0})+(N_{m,0}-N_{e,0}).$$

Now,

$$\frac{\mathrm{d}N_m}{\mathrm{d}t}=I-\alpha N_m^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(46)}$$

therefore

$$\frac{\mathrm{d}N_{e,0}}{\mathrm{d}t}+\frac{\mathrm{d}n}{\mathrm{d}t}=I_0+\frac{\partial I}{\partial h}h'-\left(\alpha_0+\frac{\partial\alpha}{\partial h}h'\right)(N_{e,0}+n)^2$$

and

$$\frac{\mathrm{d}n}{\mathrm{d}t}=-\alpha_0(N_m-N_{e,0})n+\frac{\partial}{\partial h}(I-\alpha N_m^2)h'.$$

This relation is precise but to the first order we may write

$$\frac{\mathrm{d}n}{\mathrm{d}t}=-2\alpha Nn+\frac{\partial}{\partial h}(I-\alpha N^2)h',$$

that is,

$$\frac{\mathrm{d}n}{\mathrm{d}t}=-An+Bh'. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(47)}$$

If h' varies sinusoidally semi-diurnally, that is,

$$h'=P_2\cos 2\omega t,$$

where ω is the angular velocity of the Earth, (47) has the solution

$$n=n_0\mathrm{e}^{-At}+\frac{BP_2}{\sqrt{(4\omega^2+A^2)}}\cos\left(2\omega t-\arctan\frac{2\omega}{A}\right). \quad \ldots\ldots \quad \mathbf{(48)}$$

The first term in this expression shows that the electron density recovers exponentially, with a relaxation time $\tau=1/2\alpha N$, from any initially imposed perturbation n_0.

The second term,

$$n=\frac{BP_2}{\sqrt{(4\omega^2+4\alpha^2N^2)}}\cos\left(2\omega t-\arctan\frac{\omega}{\alpha N}\right). \quad \ldots\ldots \quad (49)$$

gives the tidal perturbation of the electron density.

Now, as in fact equation (**49**) implies, because of the intrinsic stability of an ionospheric layer in which ionization and recombination processes are rapid, we should expect appreciable lunar perturbation of the F_2 only during the late afternoon and evening when the relaxation time is reasonably long, yet the E region is still the most highly conducting region of the ionosphere. This is in fact the case at Canberra, as will be shown in a later paper, and it seems reasonable to assume that at all moderate latitude stations the lunar variation extracted from data averaged over all solar hours is dominated by the behaviour at these times.

Recombination is then the dominant process so that we may write

$$B=-\frac{\partial}{\partial h}(\alpha N^2). \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (50)$$

At the height of the F_2 region at moderate latitudes (250–300 km) the recombination coefficient is proportional to the air density, that is,

$$\alpha=\alpha_0 e^{-h/H},$$

where H is the scale height,

$$B=\frac{\alpha}{H}N^2 \qquad \left(\frac{\partial N}{\partial h}\text{ is zero at the density maximum}\right),$$

and

$$n=\frac{\alpha N^2 P_2}{H\sqrt{(4\omega^2+4\alpha^2N^2)}}\cos\left(2\omega t-\arctan\frac{\omega}{\alpha N}\right). \quad \ldots\ldots \quad (51)$$

Now at night the relaxation time $(1/2\alpha N)$ of the ionosphere can be considered infinite, as its value at any time t after sunset is always greater than t. Hence the variation of critical frequency should lag that of height by 3 hr.

We may estimate the amplitude of the critical frequency (f_0) variation from (**51**)

$$\frac{n}{N}=\frac{\alpha NP_2}{2H\sqrt{(\omega^2+\alpha^2N^2)}},$$

therefore

$$\frac{\Delta f_0}{f_0}=\frac{\alpha NP_2}{2H\sqrt{(\omega^2+\alpha^2N^2)}},$$

where Δf_0 is the total variation of critical frequency, i.e. twice the amplitude of the variation.

Put $\alpha=3\times10^{-10}$ cm³/sec (Mitra),
$N=2{\cdot}5\times10^5/\text{cm}^3$,
$P_2=1{\cdot}5\times10^5$ cm,
$\omega=2\pi/(24\times3600)\ \text{sec}^{-1}$,
$H=3\times10^6$ cm.

Then

$$\Delta f_0/f_0=1{\cdot}8\times10^{-2}.$$

Martyn found a variation of about 2 per cent. at Canberra, with maximum critical frequency at 09 lunar hours, in good agreement with these results, and other non-equatorial stations show variations of similar amplitude and phase (Table 1).

The arguments above are not applicable to the F_2 region near the magnetic equator, for here, probably as a result of solar tides (Martyn 1955), the day-time F_2 region occurs at the great height of from 350 to 450 km (Maeda 1955). The recombination coefficient at this height is small and thus the electron relaxation time in the F_2 region near the magnetic equator is of the order of hours during the day when the conductivity of the E region is greatest. Vertical diffusion is inhibited by the geomagnetic field. This will lead to large day-time lunar variations of $h^{\max.}F_2$ and f_0F_2, and we should expect these day-time lunar variations to dominate the variations found in data meaned over all 24 solar hours. Figures 1 and 2 show that this is the case.

There is good evidence that at 300 km electron decay is primarily due to attachment and hence proportional to the pressure, but the rate of electron loss cannot fall indefinitely with height as attachment must ultimately cease to be the most important process of electron loss. At 400 km, then, the effective recombination coefficient has probably fallen to a constant low value.

On the other hand 400 km is well above the height of maximum ion production so that the ionization rate will be practically proportional to the pressure, that is,

$$I=I_0e^{-h/H}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (52)$$

For day-time equatorial conditions, therefore, $\partial(I-\alpha N^2)/\partial h$ becomes $-I/H$, $\partial N/\partial h$ being zero at the electron density maximum. Substituting this value in (49) we obtain

$$n=-\frac{IP_2}{H\sqrt{(4\omega^2+4\alpha^2N^2)}}\cos\left(2\omega t-\arctan\frac{\omega}{\alpha N}\right),$$

hence

$$\frac{\Delta f_0}{f_0}=\frac{IP_2}{2HN\sqrt{(\omega^2+\alpha^2N^2)}}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (53)$$

From this we may estimate the expected phase and amplitude of the Huancayo lunar critical frequency variation.

Put $N=10^6$ cm^{-2}, $\alpha=5\times10^{-11}$ cm^3/sec. Then the relaxation time $(1/2\alpha N)$ equals about 2·8 hr. This is much less than the time for which the sun shines (12 hr) so we should expect the electron density to be approximately equal to the equilibrium value and thus

$$I\simeq\alpha N^2=50\ \text{cm}^{-3}\text{sec}^{-1},$$

also, $P_2=5\times10^5$ cm, $H=3\times10^6$ cm, and $\omega=2\pi/(24/3600)$ sec^{-1}.

Substituting these values in (**53**) gives arctan $\omega/\alpha N$ equal to 55°, i.e. 1·9 hr, and $\Delta f_0/f_0$ equal to $4\cdot7\times10^{-2}$.

Thus, we should expect a lunar critical frequency variation of about 5 per cent. at Huancayo, and maximum critical frequency should occur about 4 hr before the time of maximum height. The phases and amplitudes observed by Martyn (Table 1) are in agreement with these conclusions.

IV. ACKNOWLEDGMENTS

This work forms part of the programme of the Radio Research Board of the C.S.I.R.O., and is published by permission of the Board. It was undertaken at the suggestion, and under the very helpful guidance, of the Chief Scientific Officer, Dr. D. F. Martyn.

V. REFERENCES

APPLETON, E. V., and BEYNON, W. J. G. (1948).—*Nature* **162**: 486.
BAKER, W. G. (1953).—*Phil. Trans.* A **246**: 295.
BAKER, W. G., and MARTYN, D. F. (1952).—*Nature* **170**: 1090.
BAKER, W. G., and MARTYN, D. F. (1953).—*Phil. Trans.* A **246**: 281.
BURKARD, O. (1948).—*Terr. Magn. Atmos. Elect.* **53**: 273.
CHAPMAN, S. (1919).—*Phil. Trans.* A **218**: 1.
CHAPMAN, S., and BARTELS, J. (1940).—"Geomagnetism." (Oxford Univ. Press.)
FEJER, J. A. (1953).—*J. Atmos. Terr. Phys.* **4**: 184.
GOLD, E. (1910).—*Phil. Mag.* **19**: 26.
HIRONO, M. (1950).—*J. Geomagn. Geoelect., Kyoto* **2**: 1.
HUXLEY, L. G. H. (1952).—*Aust. J. Sci. Res.* A **5**: 10.
MAEDA, H. (1955).—*Rep. Ionosphere Res. Japan* **9**: 60.
MARTYN, D. F. (1947).—*Proc. Roy. Soc.* A **190**: 273.
MARTYN, D. F. (1948*a*).—*Proc. Roy. Soc.* A **194**: 429.
MARTYN, D. F. (1948*b*).—*Nature* **162**: 142.
MARTYN, D. F. (1953).—*Phil. Trans.* A **246**: 306.
MARTYN, D. F. (1955).—"The Physics of the Ionosphere." (The Physical Society: London.)
MATSUSHITA, S. (1949).—*J. Geomagn. Geoelect., Kyoto* **1**: 17.
MITRA, S. K. (1952).—"The Upper Atmosphere." p. 290. (The Asiatic Society: Calcutta.)
PEDERSEN, P. O. (1927).—"Propagation of Radio Waves etc." (Danmarks Natur Samf.: Copenhagen.)
PEKERIS, C. L. (1937).—*Proc. Roy. Soc.* A **158**: 650.
ROCKET PANEL (1952).—*Phys. Rev.* **88**: 1027.
SCHUSTER, A. (1908).—*Phil. Trans.* A **280**: 163.
STEWART, B. (1882).—Terrestrial magnetism. *In* "Encyclopaedia Britannica." 9th Ed.

MEASUREMENTS OF COUNTER-GRADIENT HEAT FLOWS IN THE ATMOSPHERE*

By A. F. BUNKER†

[*Manuscript received November* 7, 1955]

Summary

Heat flows and temperature gradients were measured in a stable atmosphere heated by the water during a United States east coast to Bermuda aeroplane flight. The flux was computed from $c_p\bar{\rho}\overline{w'T'}$, where w' and T' are the vertical velocities and temperature deviations of the individual gusts. It is found that the heat flow is directed up the gradient as predicted by Priestley (1954). The expression for the maximum heat flow to be expected derived by Priestley in terms of the standard deviation of the temperature fluctuations and the temperature gradient gives the correct order of magnitude for the observed flow.

I. INTRODUCTION

The generally accepted theory of turbulent diffusion of momentum, heat, and water vapour was modelled after the molecular diffusion theory and considers the transport to be accomplished by turbulent eddies which carry the average properties of the fluid at the level of origin through some reference level. Whenever a gradient of the properties exists a net turbulent transport of the property results from the summation of all the eddies passing through the surface, provided there is no mass transport through the surface. This theory shows that the flux is proportional to both the gradient and the coefficient of turbulent mass exchange and the flux is directed down the gradient. It is concerning the point of the necessity for heat flow to be down the gradient even in a gravitational field that the theory fails to describe the actual flow of heat through the atmosphere. It has been observed that in certain cases upward flows of heat exist in the face of potential temperatures increasing with height.

The idea of warmer parcels of air rising through the atmosphere and contributing to a counter-gradient flow of heat was incorporated into the classical diffusion theory by Priestley and Swinbank (1947). They differed in their approach by considering that, at the point of origin of the eddy, the air did not necessarily have the mean temperature of that level but had the temperature T_0+T'', where T_0 is the average temperature of the level and T'' is a deviation from that temperature. Treating this temperature sum in much the same manner as the authors of the classical mixing length theory, they develop the heat flow equation

$$F=\rho_0c_p[-K(\partial T_0/\partial z+\Gamma)+\overline{wT''}]. \quad \ldots\ldots\ldots\ldots\ (1)$$

---

* Contribution No. 682 from the Woods Hole Oceanographic Institution. Work supported by the Office of Naval Research, U.S. Navy, under contract number Nonr 1721 (00) (NR–082–021).

† Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, U.S.A.

Here ρ_0 is the density corresponding to T_0, K is the diffusion coefficient, Γ is the dry adiabatic lapse rate, and w is the vertical velocity of the eddy at the reference level. The authors point out that the $\overline{wT''}$ term results in a systematic upward flux of heat irrespective of the sign of $(\partial T_0/\partial z+\Gamma)$. This equation cannot be checked directly because of the impossibility of observing T''. Because of this difficulty the equation is not quantitatively useful and remained without further development although it was recognized as a great advance toward a true description of heat flow in the atmosphere.

Following the presentation of the Priestley-Swinbank theory, several studies of the properties of buoyant parcels moving through an environmental air were made. One of these by Houghton and Cramer (1951) studied the effects of entrainment of environmental air upon columns of air rising as a result of its buoyancy. Equations are derived that give mass entrainment, temperature deviations, and velocities as functions of the height. These equations were set up for both dry air and saturated air from which the water vapour condenses as the air rises. Thus, assuming that the only interaction with the environment comes from the entrainment of air which carries with it its characteristic momentum and temperature anomaly, the life history of a warm parcel of air can be determined. This approach to the study of buoyant parcels was developed further by Bunker (1953), who followed the general development of Houghton and Cramer but added the effects of a turbulent atmospheric environment and a dynamic pressure drag upon the rising bubble. Equations are found which give temperature deviations and vertical velocities as functions of height when the turbulent mass exchange coefficients of the environmental air and the size of the parcels are known or assumed as functions of the height. Numerical integration of the equations is required in both of these treatments.

A theory presented by Scorer and Ludlam (1953) described a mechanism of convective rise of bubbles which erode as they penetrate the atmosphere. This mechanism is modelled after the tank experiments of Davies and Taylor (1950) and experiments of the authors with bubbles rising through tinted water. Particular emphasis is placed upon the effects of the wake of the bubble in aiding subsequent parcels to penetrate to greater heights of the atmosphere.

An extremely valuable treatment of the problem of buoyant parcels was developed by Priestley (1953). Time derivative equations of the vertical velocity and temperature excess T' of the parcel following the mean motion are developed taking into account mixing of momentum and heat of the parcel air with the environment, which has the temperature T_e. The simultaneous equations are:

$$\dot{w}=(g/T_e)T'-k_1w, \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

$$\dot{T}'=-w(\partial T_e/\partial z+\Gamma)-k_2T'. \quad\ldots\ldots\ldots\ldots\ldots \quad (3)$$

Here k_1 and k_2 are mixing rates of momentum and heat. They are defined in terms of the radius R of the parcel and the diffusivity coefficient K by the following relations, where c_1 and c_2 depend upon the form of the parcel:

$$k_1=c_1K_1/R^2, \qquad k_2=c_2K_2/R^2. \quad\ldots\ldots\ldots\ldots\ldots \quad (4)$$

The second order differential equation was formed

$$\ddot{w}+(k_1+k_2)\dot{w}+\varkappa^2 w=0 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

and solved in conjunction with the previous equations for w and T'/T_e as functions of time and the mixing rates. When these solutions are examined for various combinations of environmental lapse rates and mixing rates it is clear that the parcels may have three modes of motion : oscillatory, asymptotic, and absolute buoyancy.

In a subsequent paper, Priestley (1954) has developed the equations describing the motion and temperature relations of a buoyant parcel to the point where heat flows can be calculated under given conditions. Using either the oscillatory motion or the asymptotic motion case, the products of wT' and T'^2 are averaged over the path of the parcel to get the heat flow and the standard deviation (σ_T) of the temperature. The relation is found

$$F_H/\rho c_p=\sigma_T^2(g/T_e)\frac{2k_1}{(g/T_e)(\partial T_e/\partial z+\Gamma)+k_1k_2+2{k_1}^2}, \quad \ldots\ldots \quad (6)$$

which gives the heat flow in terms of the temperature fluctuations, the lapse rate, and the mixing rates.

To investigate the dependence of the convective heat flow as a function of lapse rate and other quantities, several parameters are defined and substituted into the above equation. From a plot of these parameters it is apparent that a maximum value of the heat flow exists for a given mixing rate ratio. This is interpreted to mean that an optimum size of parcel exists which can transport a maximum amount of heat for a given degree of turbulence and thermal stability. The upper limit for the heat flux can now be expressed as

$$F_H=\frac{\rho c_p \sigma_T^2}{\sqrt{\{(3T_e/g)(\partial T_e/\partial z+\Gamma)\}}}, \quad \ldots\ldots\ldots\ldots\ldots \quad (7)$$

when the reasonable assumption is made that $k_1=k_2$.

In view of these theories set forth by the various workers, which have had little confirmation or checking with observations, it is desired to obtain data that can be used for this purpose. The flow of cool air over warm water offers circumstances ideal for the study of the gravitational effect upon heat flow and the application of the concepts and equations of Priestley. Usually in these cases, the turbulence in the lowest 50 m is too small to transport all of the heat crossing the air-water interface, with the result that a large super-adiabatic lapse rate is established. The resulting instability allows warm buoyant parcels of varying sizes to rise through the air, thereby transporting heat upward at a faster rate than that of turbulent diffusion. This system carries the heat to higher levels where it is finally mixed into the environmental air. As a result of the small heat diffusion in the lowest levels and the greater transport to higher levels, a potential temperature minimum can be established in the 50–300 m level in the cool air. As a temperature inversion usually exists at the top of the mixed layer the stability of the air is increased further by downward diffusion of heat from this source. Under the conditions just described and at altitudes up to the inversion many of the various types of heat flow will be large and easily measured.

The major technique employed in this study is the measurement of heat flow from an aeroplane by averaging the cross products of the temperature and vertical velocity deviations. Such values have been obtained at numerous heights and localities where counter-gradient fluxes are suspected to exist. Observations were obtained from an aeroplane flying at levels from 30 to 1700 m and to distances up to 1100 km off shore. Radiosonde and radiowinds observed at land, ship, and island stations were used as supplementary sources of information. The instrumentation of the aeroplane and the technique of obtaining vertical and horizontal turbulent fluctuations of the air velocity have been described by Bunker (1955), hence only the briefest mention of these topics will be made. A modification of the reduction system is used now which gives reliable values over flight paths of many kilometres in length. This change involves the use of deviations of the airspeed and attitude from a mean over the entire run rather than over 10 sec. The fluctuations are determined from the measured responses of a vertically mounted accelerometer, a gyroscope, and a pressure gauge connected to a pitot-static tube. Temperature fluctuations are found from the response of a thermopile. Once a time series of w_i, the vertical turbulent component of the wind, and T_i, the temperature fluctuations, read to $0 \cdot 1$ °C are obtained, then cross products and root mean square deviations can be formed, giving $(\overline{w'^2})^{\frac{1}{2}}$, $(\overline{T'^2})^{\frac{1}{2}}$, and $c_p\bar{\rho}\overline{w'T'}$, the heat flux. Particular attention will be paid here to the heat flow calculation, and an interpretation of this observed quantity in terms of diffusion and buoyant transport will be discussed. Heat flows into the air mass averaged over all or part of the distance covered by the air in its travel over the water surface have been calculated from observed individual temperature changes, local changes, and mean vertical velocities in the air mass. From the mean temperatures observed at various heights by the aeroplane, the gradients of the potential temperature can be found and the way is clear for a discussion of Priestley's maximum buoyant heat flow and associated problems of turbulent mass exchange.

II. Observational Data Obtained during the Polar Outbreak over the Atlantic Ocean, January 18, 1955, Rhode Island to Bermuda

This flight to Bermuda was made at about the middle of a 3-week period of predominant north-westerly flow off the east coast of North America. A low pressure system passed to the north of New England on the morning of January 17 and north-west winds prevailed until a coastal storm interrupted them temporarily on the afternoon of the 19th. The skies were clear over the land at take-off time and remained so as the air passed over the shallow coastal waters. As warmer waters were reached, a few cumulus clouds appeared and rapidly built up to a complete overcast of stratocumulus as the air flowed over the very warm waters of the Gulf Stream. In this region the aeroplane was flying in and out of scud at 500 m. No showers were noted at any time. Some 300 km beyond the Gulf Stream breaks in the overcast appeared and grew until only six-tenths of the sky was covered by cumulus clouds at Bermuda.

The resistances of a thermistor mounted in a housing attached ahead of the nose of the aeroplane were recorded continuously throughout the flight. From these resistances the air temperatures were found, corrected for the dynamic heating of the aeroplane's motion, and potential temperatures computed. Water temperatures were obtained from mean maps and the intake water temperatures of the R.V. *Atlantis* which travelled the course to Bermuda January 19–22, 1955. These water temperatures are treated as air temperatures and reduced to potential temperatures. This procedure is justified on the basis that the air immediately in contact with the water takes on the water's temperature.

Five times *en route* to Bermuda the aeroplane was flown from the cloud base down to 90 m above the water to obtain vertical temperature gradients, turbulence, and heat flow data. The temperatures obtained during these descents had to be corrected for the horizontal gradients since 30–40 km was traversed during a descent and ascent. The temperatures, corrected to mean off-shore distances, have been plotted on a height-temperature diagram, and presented in Figure 1. As conclusions of considerable importance will be drawn from the nature of these gradients great care was taken that no systematic errors were introduced by the reduction system and the flight procedure of flying faster at the lowest elevations. From numerous flight tests performed it is concluded that if any systematic errors exist they are less than 0·1 °C. An uncertainty of 0·1 °C over a 500 m height range is equivalent to an uncertainty of 2×10^{-6} cm^{-1} in the gradient determination. The observed gradients are several times this amount. It is noted that a slight but real stability exists in the upper 90 per cent. of the sub-cloud layer. Although no temperatures other than the water temperatures were obtained below the 30 m level it is obvious that the air in this region is very unstable.

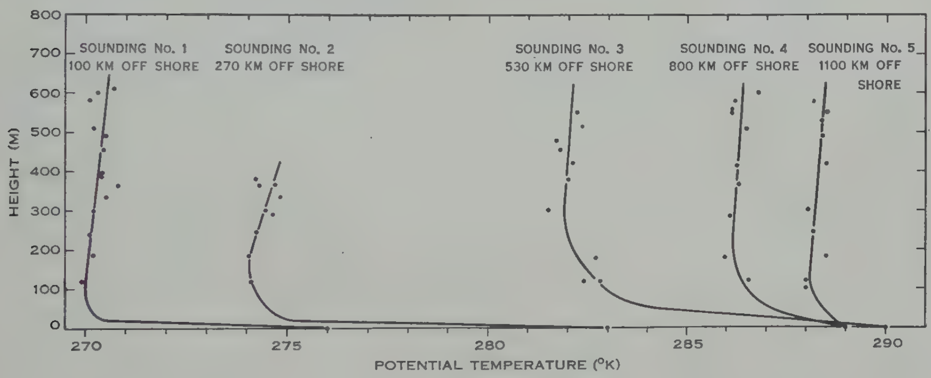


Fig. 1.—Potential temperature gradients, Rhode Island to Bermuda.

Fluctuation data obtained from 17 horizontal runs made during the five descents already mentioned were reduced to time series of w_i and T_i and subsequently to series of w' and T'. From these series values of $(\overline{w'})^{\frac{1}{2}}$, $(\overline{T'})^{\frac{1}{2}}$, and $c_p\rho\overline{w'T'}$ have been computed and tabulated in Table 1. With these values are listed information concerning an identifying run number, distance off shore, height above the sea surface, and the vertical gradients of the potential temperature measured from the temperature-height diagram, Figure 1. The last column gives the maximum heat flow as computed from Priestley's theory. This will be discussed later.

A potential temperature cross section of the atmosphere between Rhode Island and Bermuda, presented here as Figure 2, has been drawn to show the temperature changes taking place in the air mass as it moves over the water and to show other observed quantities. The ordinate of the diagram is atmospheric pressure in millibars, while the abscissa is the double entry of the Greenwich Mean Time of a particular observation made from the aeroplane, and of the off shore distance from the Rhode Island coast. The heavy lines are isopleths of potential temperature computed from both aeroplane measurements, represented here by dots, and radiosonde measurements, represented by crosses within circles. Cloud reports are entered along the top and at the flight level. The potential temperature of the air in contact with the water is entered in the ellipses at the sea-level pressure line. The turbulence encountered during the flight is indicated by the root-mean-square vertical velocities in cm sec^{-1} entered in the boxes located at the proper pressure and time. Lastly the heat flows in mcal cm^{-2} sec^{-1} computed from the cross products of the vertical velocities and the temperature deviations are entered with an arrow representing the direction of flow.

TABLE 1

HEAT FLOWS COMPUTED FROM FLUCTUATION DATA, JANUARY 18, 1955

Atlantic Ocean, Rhode Island to Bermuda. 1820–2210 hr G.M.T.

Run Number	Distance (km)	Height (m)	$(\overline{w'^2})^{\frac{1}{2}}$ (cm sec^{-1})	$(\overline{T'^2})^{\frac{1}{2}}$ (°C)	$c_p\rho\overline{w'T'}$ (mcal cm^{-2} sec^{-1})	$d\theta/dz$ (cm$^{-1}\times 10^5$)	$F_{H\max.}$ (mcal cm^{-2} sec^{-1})
974	100	520	86	0·05	0·1	1·1	0·2
975	100	335	121	0·07	1·7	1·1	0·4
976	100	150	94	0·08	1·1	1·1	0·6
977	270	305	107	0·08	1·1	3·0	0·4
978	270	150	135	0·19	3·9	−1·0	—
979	530	460	174	0·10	2·1	0·5	1·3
980	530	120	177	0·18	1·3	−8·0	—
981	800	550	122	0·09	1·1	0·6	1·0
982	800	305	134	0·10	2·0	0·6	1·2
983	800	120	138	0·11	2·1	−7·0	—
984	1100	550	129	0·07	0·8	0·8	0·5
987	1100	305	151	0·08	1·1	0·8	0·7
990	1100	90	138	0·13	2·4	?	—

The sensible heat flowing into the moving air column has been calculated by measuring the temperature change per 100 km as observed from the aeroplane and by measuring local changes and subsidence rates by use of radiosonde and

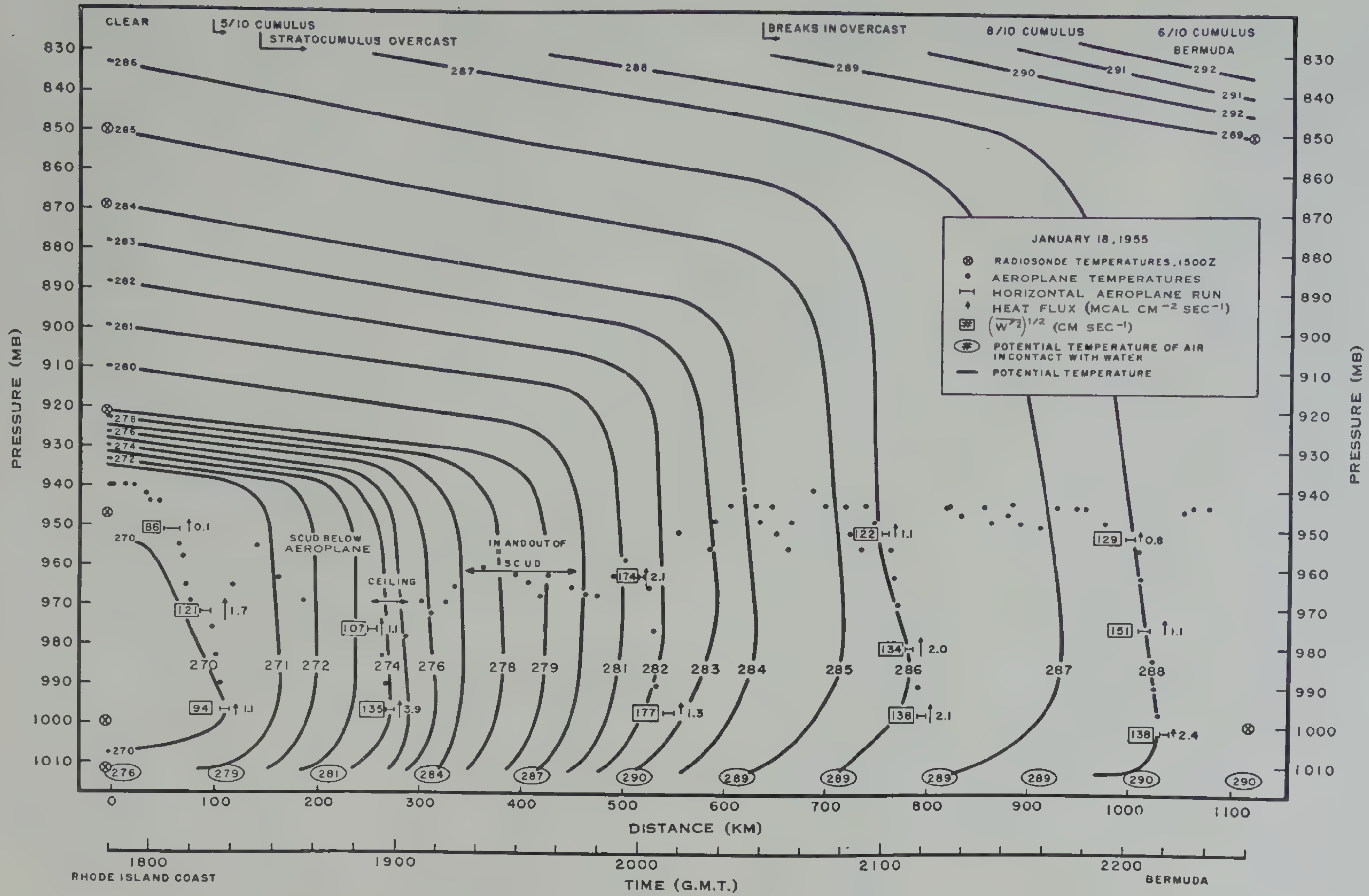


Fig. 2.—Atmospheric cross section from Rhode Island to Bermuda.

radiowind reports. The temperature changes were measured in two height ranges, one below the inversion and one in the inversion. From these figures the amount of heat added was found and the rate of addition calculated by dividing by the time of transit, 8×10^3 sec. Adding the rates of heating for the two regions gives the total heating assuming no subsidence and no local changes of temperature. The average over the entire area was found to be $5 \cdot 3$ mcal cm^{-2} sec^{-1}. To evaluate the heating arising from the neglected effects the terms in the following equation were determined

$$\left.\frac{\mathrm{d}Q}{\mathrm{d}t}\right)_{l\,\&\,s} = \left(\frac{\mathrm{d}p}{g}\right) c_p \left[\frac{\mathrm{d}T}{\mathrm{d}t} + \left(\frac{p}{10^6}\right)^{R/c_p} w \frac{\mathrm{d}\theta}{\mathrm{d}z}\right], \quad \ldots\ldots\ldots\ldots \quad (8)$$

where $\mathrm{d}Q/\mathrm{d}t)_{l\,\&\,s}$ is the heating due to subsidence and local changes in temperature where $\mathrm{d}p$ is the change in pressure in baryes through the layer concerned. The $\mathrm{d}T/\mathrm{d}t$ term was measured by using the Bermuda radiosonde for two times 12 hr apart. The vertical velocity, w, was computed from the radiowinds from Nantucket, Mass., Washington, D.C., and Bermuda, by computing mass divergences. These divergences were calculated by the Bellamy (1949) method. Assuming the vertical velocity to be zero at the 1000 mb level, it is found that the air is subsiding at the rate of $0 \cdot 6$ cm sec^{-1} at 850 mb over the entire region. While it is admitted that the computation of divergences and velocities from so few wind observations is generally inaccurate, in the present case the wind field was steady for long periods and over large areas, so the computed velocity is considered reasonably reliable. The average rate of heating due to these two terms is $1 \cdot 3$ mcal cm^{-2} sec^{-1} in the 1000–800 mb layer. When this is subtracted from the average total heat flow measured for the same layer, we get $4 \cdot 0$ mcal cm^{-2} sec^{-1}. The maximum heat flow found for the 100 km regions was found to be $8 \cdot 0$ mcal cm^{-2} sec^{-1} for the 200–300 km region. It is of interest to compare these values with those entered in Table 1. The average for the total area, $4 \cdot 0$ mcal cm^{-2} sec^{-1}, represents satisfactory agreement, all difficulties considered, with $2 \cdot 4$ mcal cm^{-2} sec^{-1}, the average of the five cross product heat flows taken at the lowest level of a given series. Likewise the heat flow for the 200–300 km region computed from the change in heat content, $8 \cdot 0$ mcal cm^{-2} sec^{-1}, shows satisfactory agreement with the cross products method value of $3 \cdot 9$ mcal cm^{-2} sec^{-1}.

III. Description of Convective Parcels and Environment

Before checking the heat flows computed according to any theory against the observed heat flows, a few values and averages have been assembled that will give a clearer conception of the convective regime existing in the study area. First it should be pointed out that at the levels at which the aeroplane was flying it is easy to pick out the individual buoyant parcels from the environmental air. Visual inspection of the horizontal w', T' time series shows clearly that the environmental air is relatively quiet and homogeneous and that the convective parcels break through this quiet air as isolated well-defined turbulent disturbances of warmer updrafts and colder downdrafts. This regime contrasts strongly with a wind field characterized by mechanical turbulence in which no warm updrafts or cool downdrafts can be recognized although w', T' cross products show that an appreciable heat flow exists. In the present study, the existence of these parcels is much more obvious at the lower levels (150–300 m) where the temperature fluctuations are greater, than at the higher levels (600 m). At the higher levels the cool downdrafts disappear so that only the environmental air and the warm updrafts are observable by inspection.

TABLE 2

OBSERVED SIZES OF COOL AND WARM AIR PARCELS

Size Range (m)	Number of Cool Air Parcels	Number of Warm Air Parcels
50	6	7
50–100	7	7
100–200	3	6
200–300	1	6
300–400	0	3
400–500	0	1

Three temperature deviation groups have been defined by $T' \leqslant -0 \cdot 2$ °C, $-0 \cdot 1$ °C$\leqslant T' \leqslant 0 \cdot 1$ °C, and $0 \cdot 2$ °C$\leqslant T'$, to aid in describing the observed parcels. These groups are intended to represent cool descending parcels, environmental air, and warm rising parcels respectively. The diameters of these cool and warm parcels have been determined for all 13 horizontal runs made during the flight to Bermuda. During these runs which totalled about 40 km only 47 parcels were observed. Their sizes are tabulated in Table 2.

TABLE 3

ANALYSIS OF TEMPERATURE CLASS PROPERTIES

Run Number	Temperature Class (°C)	Number of 20 m Intervals	Average Temperature (°C)	Average Vertical Velocity (cm sec^{-1})	Heat Flow (mcal cm^{-2} sec^{-1})
977	$-0 \cdot 2 \leqslant T \leqslant 0 \cdot 3$	165	0·006	− 4·9	1·13
977	$0 \cdot 2 \leqslant T \leqslant 0 \cdot 3$	45	0·11	15	0·41
977	$-0 \cdot 2 = T$	15	−0·18	−60	0·25
977	$-0 \cdot 1 \leqslant T \leqslant 0 \cdot 1$	105	−0·013	− 5·5	0·50
978	$-0 \cdot 3 \leqslant T \leqslant 0 \cdot 6$	165	0·004	0·6	3·92
978	$0 \cdot 2 \leqslant T \leqslant 0 \cdot 6$	28	0·346	162	2·88
978	$-0 \cdot 3 \leqslant T \leqslant 0 \cdot 2$	9	−0·244	−96	0·36
978	$-0 \cdot 1 \leqslant T \leqslant 0 \cdot 1$	128	−0·053	−28	0·71
982	$-0 \cdot 2 \leqslant T \leqslant 0 \cdot 3$	165	0·002	2·0	1·99
982	$0 \cdot 2 \leqslant T \leqslant 0 \cdot 3$	15	0·25	186	1·15
982	$-0 \cdot 2 = T$	2	−0·20	−64	0·08
982	$-0 \cdot 1 \leqslant T \leqslant 0 \cdot 1$	148	−0·02	−16	0·80

It should be noted in interpreting these size distributions that parcels larger than 1000 m would be difficult to recognize by the method used. This is true since the defined temperature groups are with respect to a line of regression of temperature upon time along the 3000 m flight path. Also it is likely that parcels of greater size have smaller amplitudes of the temperature deviations and hence may be unobservable with the present thermopile. In any case these parcels contribute the major portion of the heat flows tabulated in Table 1.

The sensible heat transported by air parcels belonging to each of the defined temperature classes has been computed for three horizontal runs picked at random from the 13 available. For each run and class, the number of 20 m intervals of the run occupied by each class, the average temperature deviation, the average vertical velocity, and the heat flow computed from the w', T' cross product have been tabulated in Table 3.

This collection of data gives a clear picture of the properties of the convective parcels and their environment, but is too incomplete to serve as a check upon the merits of the several parcel theories already discussed.

IV. Comparison of Theories of Heat Flow with the Data

A study of Figure 1 shows that the gradient of the potential temperature, except for the lowest 100 m or so, was positive throughout all of its journey from the mainland to Bermuda in spite of heating from the bottom. The discussion of systematic errors in the previous section shows that the gradients are real and that the classical heat flow theory can break down as has been pointed out previously by Priestley and Swinbank (1947).

The heat flow of the environmental air ($-0{\cdot}1 \leqslant T \leqslant 0{\cdot}1$) was computed with the idea in mind that the correlation between the vertical velocity deviations and the temperature deviations might be negative in accordance with the classical diffusion theory, which says that heat should flow down the gradient. The heat flow was found to be positive in all three cases, two of these cases having positive gradients, thus giving additional confirmation of the inadequacy of classical theory in the presence of a gravitational field. As already stated, it has been anticipated that if a temperature class with sufficiently small limits had been analysed a negative correlation would have been found. This was based on the concept that this temperature class represented the undisturbed environmental air and that for small deviations neither organized overturning nor the weak buoyancy forces of Priestley and Swinbank would cause the $w'T''$ term of (**1**) to predominate over the diffusion term. As the thermopile could not resolve temperature variations of less than 0·1 °C, no answer could be obtained to this question.

The present set of observations is not well adapted to shed light upon the validity or weaknesses of the three parcel theories of Houghton and Cramer, Bunker, and Scorer and Ludlum in their present form. No definitive statistical study of the sizes, velocities, and temperature amplitudes at various heights and distances off shore has been attempted. This was not done because, after a brief inspection of the records, it was evident that a very large range existed in these variables and that a very small sample had been taken in each location. It was felt that no conclusions could be drawn concerning the validity of the equations from this type of data without further development of the equations on a statistical basis.

As Priestley's equation (**7**) for the heat flow is expressed in observable statistical parameters, a comparison between his predicted maximum heat flows and $c_p\bar{\rho}\overline{w'T'}$ can be made. It is assumed that in the present case the heat flow mechanism of the atmosphere became adjusted in such a manner that a close approximation to maximum flow is attained. The maximum heat flow has been computed from (**7**) by substituting the observed value of $\overline{T'^2}$ for σ_T^2 as defined. In substituting $\overline{T'^2}$ for σ_T^2, one is replacing σ_T^2 by a value that is a time-weighted average of the temperature deviations of both parcels moving through the flight path and the environmental air. Equation (**7**) then gives the time-weighted maximum heat flow rather than the heat flow due to a given parcel of the optimum size. The value obtained in this manner is equivalent to the heat flow determined from $c_p\bar{\rho}\overline{w'T'}$. These values have been tabulated in the last column of Table 1. A glance will show that these are values of the same order of magnitude but smaller than the values of $c_p\bar{\rho}\overline{w'T'}$ listed in column 6 of the table.

The agreement as to sign and order of magnitude proves that the theory is sound and that the underlying assumptions must be basically correct. The fact that the predicted values are smaller than the observed values may be caused by discrepancies between the assumptions made concerning impressed temperatures in a stable atmosphere and the condition of the temperatures and vertical velocities of the buoyant parcels obtaining at the base of the stable region of the atmosphere. If this is the case then a different expression for the original conditions of the parcel might be developed that would fit the observed conditions even better. Similarly, a more carefully controlled experiment might fit the conditions treated by Priestley more exactly and lead to a closer agreement.

V. Acknowledgments

The author wishes to acknowledge the help of Mr. J. G. Fraser in servicing the instruments used in the aircraft and in operating many of them during the observational flight. The author is indebted also to Mrs. Martha Vaccaro for reading and reducing the vertical velocity and temperature data and computing the heat flows.

VI. References

Bellamy, J. C. (1949).—*Bull. Amer. Met. Soc.* **30** : 45–9.
Bunker, A. F. (1953).—*J. Met.* **10** : 212–18.
Bunker, A. F. (1955).—*J. Met.* **12** : 445–55.
Davies, R. M., and Taylor, G. I. (1950).—*Proc. Roy. Soc.* A **200** : 375–90.
Houghton, H. G., and Cramer, H. E. (1951).—*J. Met.* **8** : 95–102.
Priestley, C. H. B. (1953).—*Aust. J. Phys.* **6** : 279–90.
Priestley, C. H. B. (1954).—*Aust. J. Phys.* **7** : 202–19.
Priestley, C. H. B., and Swinbank, W. C. (1947).—*Proc. Roy. Soc.* A **189** : 543–61.
Scorer, R. S., and Ludlum, F. H. (1953).—*Quart. J. R. Met. Soc.* **79** : 94–103.

SHORT COMMUNICATIONS

ON THE DIPOLE RESONANT MODE OF AN IONIZED GAS COLUMN*†

By G. H. KEITEL‡

In a recent paper, Makinson and Slade (1954) examined theoretically the scattering of a transverse-electric radio wave by an ionized gas column. In their treatment, a smooth radial distribution of ion density was replaced by a stepped approximation and the quasi-static equations (Kaiser and Closs 1952) were used to consider scattering by this stepped column. They claim that there exists a multiplicity of resonant dipole modes for a small cylindrical column in which the radial distribution of ionization is described by a Gaussian function. Previously, Kaiser and Closs (1952) had found a single, broad resonance peak by an approximate numerical solution of the quasi-static equations, using the Gaussian distribution itself. The purpose of this note is to suggest that the multiple resonances found by Makinson and Slade were created by the discontinuities in their approximation to the Gaussian function, and that the actual Gaussian distribution exhibits only one resonance.

For transverse polarization of a plane radio wave normally incident on a small column of ionized gas, the dipole mode (cos θ variation of the scattered far fields around the column), is most significant. The scattered fields may be considerably greater than predicted by the Born approximation (which assumes that the incident wave is not attenuated appreciably in passing through the column), due to the effect of charge separations within the column. The charges separate whenever there is an electric field across a volume in which the density of ionization is not constant. The term " resonance " is used here to indicate that the transverse scattering coefficient exceeds the parallel scattering coefficient.

On the basis of a five-step approximation, Makinson and Slade found that the column will exhibit five resonant dipole modes, at values of $\lambda=1\cdot17$, $1\cdot83$, $3\cdot44$, $8\cdot73$, and $29\cdot5$, where $(1-\lambda)$ is the relative dielectric constant at the column axis. The authors claim that " there is no reason to believe that the stepped distribution manufactures resonances with no counterparts for the smooth distribution ".

However, there is sufficient evidence to cast grave doubt on these claims. Preliminary calculations using stepped distributions made by Eshleman several years ago (personal communication) showed that a plasma resonance peak

---

* The work described herein was supported by Air Force Cambridge Research Center, Air Research and Development Command, under Contract AF 19(604)–1031.

† Manuscript received October 4, 1955.

‡ Radio Propagation Laboratory, Stanford University, Stanford, California; present address: Cavendish Laboratory, Cambridge, England.

would be obtained for each step. Each resonance appeared to be due to surface charges which build up at the discontinuities in the ionization distribution, with each surface being most effective at distinct size-wavelength ratios. These resonances were very sharply peaked. Therefore this method of approximating transverse scattering by a column of ionization with a smooth radial distribution was discarded.

The exact evaluation of reflection coefficients of a Gaussian column of ionization has been performed at Stanford University using the wave-matching technique (Keitel 1955). The differential equations which describe the electromagnetic fields within the column were numerically integrated using a high-speed digital computer, the SWAC (National Bureau of Standards Western Automatic Computer). In order to be able to integrate past the singular point of the differential equation for transverse polarization, collisional losses were included in the expression for the relative dielectric constant. Line densities of ionization of 10^{13}, 10^{15}, and 10^{17} electrons per metre were considered, for various column sizes and collisional frequencies. The results of these computations, described elsewhere (Keitel 1955), did not exhibit the multiple resonance behaviour claimed by Makinson and Slade.

Subsequently, a special series of calculations has been made for a small column of fixed size, $(2\pi r_0/\lambda_0)=0\cdot1$, for various values of λ (corresponding to the notation of Makinson and Slade). The parameter r_0 occurs in the Gaussian distribution function, $\exp[-(r/r_0)^2]$, and λ_0 is the free-space wavelength. The complex relative dielectric constant is $\{1-\lambda(1-i\delta)\exp[-(r/r_0)^2]\}$. In terms of the parameters describing the column of ionization,

$$\lambda=\frac{\Gamma q}{[2\pi r_0/\lambda_0]^2(1+\delta^2)},$$

where $\Gamma=e^2/\pi\varepsilon_0 mc^2=1\cdot1271\times10^{-14}$ m,
q=line density of ionization, electrons per metre,
$\delta=\nu/\omega$=loss ratio,
ν=collisional frequency,
ω=angular frequency of the incident wave,
e=electron charge,
m=electron mass,
ε_0=permittivity of free space.

A loss ratio δ of 10^{-4} was used since this has been found to be small enough to have a negligible effect on the scattering coefficient while making it possible to integrate past the singular point at the critical radius (i.e. the radius at which the real part of the dielectric constant is zero).

The values of λ were chosen sufficiently close together so that there could be no question as to any resonance points being missed. The results are shown in Figure 1. The curve for the parallel reflection coefficient was obtained from the zero-order mode, which is sufficient for the size trail considered. The transverse reflection coefficient curve considered only the dipole mode, the zero-order and higher order modes being insignificant for this size trail. There is only one very broad resonance centred at approximately $\lambda = 2 \cdot 2$, corresponding to a relative dielectric constant of $-1 \cdot 2$ on the axis of the column. This agrees very well with the approximate results of Kaiser and Closs. The locations of the resonances claimed by Makinson and Slade are indicated in the figure. Three of their five resonances are seen to fall in the range of the calculated resonance. However, separate resonances are not observed. It is doubtful that the transverse reflection coefficient would drop down to the parallel reflection coefficient between any two calculated points in the resonance region.

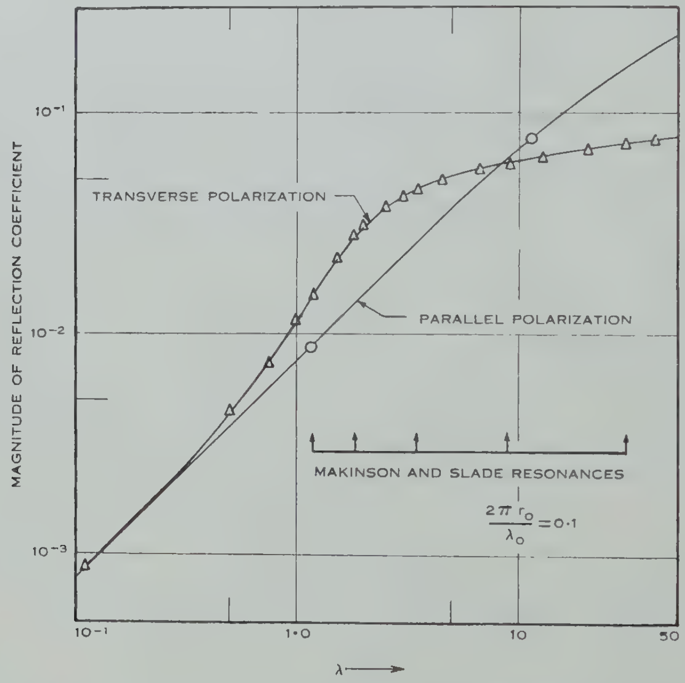


Fig. 1.—Parallel and transverse reflection coefficient of a small Gaussian column as a function of λ, where $(1-\lambda)$ = relative dielectric constant at column axis.

Makinson and Slade have based their claim of multiple resonances on their approximate solution of scattering by a stepped approximation to the Gaussian column. But this type of approximation introduces infinite gradients of electron density at each step. If one examines the differential equations which describe the internal fields, the ratio of the gradient of the dielectric constant to the value of the dielectric constant is found to be very important. For the stepped column, this ratio becomes infinite at each step because of the infinite gradient. For the smooth Gaussian distribution, this ratio becomes large only at the critical radius, where the real part of the dielectric constant is zero. Kaiser and Closs considered the gradient term for the Gaussian column in the neighbourhood of the critical radius. This author has considered this term throughout the Gaussian column. In both cases, only one resonance was found. Thus, it would appear that the multiple resonances found by Makinson and Slade *are* manufactured by the discontinuities in the stepped approximation and are not present for the smooth distribution.

The assistance of the personnel at Numerical Analysis Research, University of California at Los Angeles, in arranging for the use of the SWAC (National Bureau of Standards Western Automatic Computer) is gratefully acknowledged. Drs. V. R. Eshleman and O. G. Villard, Jr., of Stanford, were especially helpful during this research.

References

Kaiser, T. R., and Closs, R. L. (1952).—*Phil. Mag.* **43** : 1–32.

Keitel, G. H. (1955).—Mode solutions for radio waves scattered by meteor trails, Scientific Report No. 1 (AFCRC TN-55-355), Radio Propagation Laboratory, Stanford University, Stanford, California. (Condensed version published in *Proc. Inst. Radio Engrs., N.Y.* **43** : 1481–7.)

Makinson, R. E. B., and Slade, D. M. (1954).—*Aust. J. Phys.* **7** : 268–78.

NEUTRON FLUXES UNDERGROUND*

By K. B. Mather†

Eugster (1954) reported the neutron flux under 2100 m rock (5880 m water equivalent) to be 26 times that at the ground surface and suggested that radiation was coming from the centre of the Earth. An attempt has been made to confirm this interesting finding by comparing the neutron flux in the Broken Hill mines with that at ground level. Although the counting statistics are rather meagre they should be sufficient to reveal such a marked effect as claimed in the Swiss experiment.

The purpose of the present note is to report a negative result for the experiment and to draw attention to the factors which may have led to erroneous values in Eugster's experiment.

Ilford C2 boron-loaded plates, 200 μ thick, were located at various levels of the Zinc Corporation and North Broken Hill mines, controls being situated in the offices adjacent to the respective head frames of the mines. Plates were contained in sealed tins along with silica gel which served as a dehydrant to minimize the influence of fading of the latent image. Prevailing temperatures were mild at the surface ($\sim$60 °F) but increased to $\sim$85 °F at the deepest level which was 3060 ft underground (2500 m water equivalent) in the Zinc Corporation mine. Exposures lasted for approximately one month.

The plates employed were cleared of tracks by exposure to hydrogen peroxide vapour after arrival from Britain, flown to Broken Hill in cadmium-covered boxes, and returned in the same way at the conclusion of the exposure.

---

* Manuscript received October 18, 1955.

† Australian Atomic Energy Commission, Physics Department, University of Melbourne.

The short flights resulted in a small background of α-tracks due to cosmic ray neutrons in free space and this was determined and subtracted by despatching a group of identical plates over the same route under the same conditions but not permitting any other exposure. The difference should therefore be due strictly to neutrons reaching the plates when they were located on the site.

The combined results from both mines are shown in Table 1. The very low density of tracks (~20–200/cm²) made scanning very tedious and the statistics poor. Uncertainties in the values shown are based on counting statistics only. The total number of tracks recorded was 173.

TABLE 1

DENSITY OF α-TRACKS FROM THE $^{10}B(n,\alpha)^{7}Li$ REACTION DUE TO NEUTRON FLUXES UNDERGROUND

The background figure (column 3) should be subtracted from all other values

Depth of Rock below Ground (ft)		α-Tracks per cm³ (measured)	α-Tracks per cm³ (background)
NBH* & ZC	Surface (~1000 ft above sea-level)	$(0\cdot96\pm0\cdot14)\times10^4$	$(0\cdot18\pm0\cdot03)\times10^4$
NBH	1000	$(0\cdot29\pm0\cdot06)\times10^4$	
NBH	2000	$(0\cdot17\pm0\cdot03)\times10^4$	
ZC	3000	$(0\cdot24\pm0\cdot04)\times10^4$	

* NBH = North Broken Hill mine, ZC = Zinc Corporation mine.

Assuming that the α-yield from the $^{10}B(n,\alpha)^{7}Li$ reaction correctly reflects the prevailing neutron flux, it is apparent that there is little difference between the various plates with the exception of the surface exposure which is significantly higher. There is no indication of a trend towards higher neutron fluxes underground, in direct contradiction of the 26-fold increase claimed by Eugster for a depth of 6900 ft underground.

The danger exists in experiments of this kind that track fading during the long exposure will lead to erroneous absolute values, which, if conditions are not identical at the several sites, tend to make comparisons hazardous. The much longer exposure period (116 days) employed by Eugster for his surface plates would tend to depress the neutron flux estimate there, if fading effects were significant, relative to the underground value which was based on a 37-day exposure. This point has been referred to previously (Mather 1954).

In the present experiment also, fading could be present and would tend to operate in the opposite direction because, although exposure times were identical, the underground temperature was considerably higher and fading is generally accelerated by elevated temperatures. A useful check on this can be obtained by comparing the density of contamination tracks on all the plates scanned. Both the emulsion and glass of a photographic plate contain detectable thorium traces which gradually build up a background of α-stars and single tracks. If fading is also present an equilibrium track density is reached, after a time depending on the fading rate, when the generation of new tracks is exactly compensated by the loss of old ones. This contamination background was compared on the various plates with a view to comparing the prevailing fading rates. There was no significant difference, the α-track distribution being constant at a figure of about $6 \times 10^4/\text{cm}^3$ over the whole series.

While the underground measurements listed in Table 1 are rather crude and not very consistent, they appear to average somewhat higher than the background value, perhaps by about 500 α-tracks/cm³, or 16 α-tracks $\text{cm}^{-3}\ \text{day}^{-1}$. It is of interest to compare this yield with that expected from neutron generating processes which are feasible in the crust of the Earth.

First consider spontaneous fission: ^{238}U having a fission rate of $6 \cdot 9 \times 10^{-3}\ \text{g}^{-1}\ \text{sec}^{-1}$ (Segre 1952) is the most prolific isotopic contributor in any given percentage abundance. On the average 2·2 neutrons are emitted per spontaneous fission so that its source strength (neutrons $\text{cm}^{-3}\ \text{day}^{-1}$) is $3 \cdot 5 \times 10^3 f$, where f is the fractional concentration of uranium in the ground, by weight, and the density of ground has been taken as 2·7 g/cm³.

Another source of ground neutrons of comparable strength is provided by (α, n) reactions on the major elements (oxygen, silicon, and aluminium) present in the ground, the α-particles coming from radioactive inclusions in the ground. The number of α-particles released per unit time depends on the radioactive series present and their state of equilibrium. In the Broken Hill region uranium is certainly present in trace quantities. If in complete equilibrium UI emits 10^5 α-particles $\text{g}^{-1}\ \text{sec}^{-1}$, which may be taken as an upper limit for the actual emission rate in the present case. Thick target (α, n) yields have been measured for polonium 5·3 MeV α-particles, giving 0·07 neutrons/10^6 α-particles for oxygen, 0·16 for silicon, and $\sim 0 \cdot 7$ for aluminium (Segre 1953). Taking a weighted average of $0 \cdot 14/10^6$ for earth and reducing this to a half to allow for the reduced barrier penetration of lower energy uranium α-particles leads to a neutron yield of $1 \cdot 6 \times 10^3 f\ \text{cm}^{-3}\ \text{day}^{-1}$, where f is the uranium concentration as before. (However, quite small concentrations of certain other elements could alter this figure appreciably. For instance, the (α, n) yield on beryllium is $40/10^6$ U α-particles.)

Other processes which suggest themselves are: (1) (γ, n) reactions. The natural γ-ray energies from the uranium series are below the photo-neutron thresholds. (2) The μ-meson–star process releasing neutrons. While this will contribute to the exposure of plates at the surface, and in fact some distance underground, its contribution will be negligible at the great depths employed in the present experiment (see Wilson 1952). Both these processes may therefore be excluded as significant neutron sources.

Combining spontaneous fission and (α, n) yields leads to a total source strength:

$$q \simeq 5 \cdot 1 \times 10^3 f\ \text{cm}^{-3}\ \text{day}^{-1}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

If it is assumed that the ground consists of a homogeneous mixture of elements and also that no neutrons are absorbed during slowing down, an upper limit can be obtained for the corresponding neutron flux:

$$\varphi = q\Lambda, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

where Λ is the absorption mean free path $(=1/N_v\sigma_a)$ of thermal neutrons. N_v is the number of atoms/cm³ and σ_a their absorption cross section. The flux is related to $I(\alpha)$, the number of α-tracks formed in the emulsion per cm³ per day, by :

$$\varphi=I(\alpha)/N_{vB}\sigma_B, \quad \ldots\ldots\ldots\ldots \quad (3)$$

where σ_B is the effective cross section for the (n, α) reaction and N_{vB} the number of boron atoms/cm³ of emulsion $(1\cdot28\times10^{21})$. Hence

$$q=I(\alpha)/\Lambda\sigma_B N_{vB},$$

and

$$f\simeq3\cdot5\times10^{-6}I(\alpha). \quad \ldots\ldots\ldots\ldots \quad (4)$$

An average ground composition similar to that listed for the Earth's crust (Forsythe 1954) has been assumed in order to obtain a rough estimate of Λ for ground. The result is $\sim$200 g/cm².

It follows that $I(\alpha)\sim16$ would require f to be $\sim0\cdot006$ per cent. in order to account for the magnitude of the neutron flux underground. However, this figure is based on the difference between two small numbers and should not be taken to indicate better than the order of magnitude. On the other hand to account for the surface yield of α-particles would call for a uranium concentration of $\sim0\cdot1$ per cent., which certainly does not exist in the Broken Hill area.

The greater part of the exposure to neutrons at the ground surface presumably arose from cosmic radiation. Taking it to be responsible for $0\cdot7\times10^4$ α-particles/cm³, i.e. $\sim$230 α-particles cm^{-3} day^{-1}, the cosmic ray neutron flux would have to be (by equation (3)) approximately 290 neutrons cm^{-2} day^{-1} at 1000 ft altitude (995 g/cm²). This is a reasonable agreement with measurements made at a similar geomagnetic latitude in Australia several years ago (Mather 1954) which indicated $\sim$230 neutrons cm^{-2} day^{-1}.

It is a pleasure to acknowledge the courtesy shown by the staff of the Zinc Corporation Ltd. and North Broken Hill Ltd. in providing for this experiment to be conducted in their mines. The interest of Professor L. H. Martin in the project has been much appreciated.

References

Eugster, J. (1954).—*Rev. Sci. Instrum.* **25** : 5.

Forsythe, W. E. (1954).—" Smithsonian Physical Tables." 9th Ed. (Smithsonian Institution : Washington.)

Mather, K. B. (1954).—*Aust. J. Phys.* **7** : 601.

Segre, E. (1952).—*Phys. Rev.* **86** : 21.

Segre, E. (1953).—" Experimental Nuclear Physics." Vol. 2, p. 366. (J. Wiley & Sons: Inc. : New York.)

Wilson, M. A. (1952).—" Progress in Cosmic Ray Physics." p. 425. (North-Holland Publishing Co. : Amsterdam.)

THE FITTING OF A TRUNCATED LOG-NORMAL CURVE TO DAILY RAINFALL DATA*

By S. C. Das†

Introduction

In a previous paper (Das 1955) the author discussed a problem of curve fitting which arose in testing the hypothesis proposed by Bowen (1953) concerning daily rainfall data.

In this investigation, the Sydney daily rainfall over the period of 94 years from 1859 to 1952 was examined. A type III probability distribution of the form

$$f(x)=\frac{\mu^{\varkappa}}{\Gamma(\varkappa)}e^{-\mu x}x^{\varkappa-1}$$

provided a good fit to these data.

Because, in the case of rainfall data, we measure the rainfall x to the nearest rounded-off unit on some scale, we are likely to have some zero values of x; in fact we have a large number of zero values in the case of daily rainfall. This made it impossible to use the maximum likelihood equations for estimating the parameters, since these equations involve the sum of the logarithms of the observations. We therefore fitted truncated type III curves to the observations, the truncation being of the following two types. In one case we chose a small interval $(0, \delta)$ and truncated the distribution at δ, ignoring the actual values of x less than δ, but using the fact that we knew their total number. In the other case we fitted a truncated type III curve to the observations which are greater than δ, and ignored all observations which are less than δ. In both cases a very good fit was obtained as judged by the χ^2 test, and there was no significant difference in the expected numbers in the truncated part, so that there was no evidence of singularity at the origin of the distribution.

It is more usual, however, in meteorological practice to fit a log-normal curve of the type

$$f(x)=\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\cdot\frac{1}{(x-a)}\exp\left[-\frac{1}{2\sigma^2}\{\ln(x-a)-\mu\}^2\right]$$

to rainfall data.

The first attempt to fit a normal distribution to the logarithms of the values of a meteorological element appears to have been made by Blackhouse (1891). He compared the frequencies of annual rainfall amounts and of their logarithms with a normal distribution. He took $a=0$ and found the fit not very convincing, perhaps because only 30 observations (1860–89) were available. More recently

---

* Manuscript received October 31, 1955.

† Australian National University, Canberra, A.C.T.

Brooks and Carruthers (1953, p. 102) suggest that a log-normal curve will give a good fit for any distribution which is positively skew and sufficiently leptokurtic. They fitted a log-normal curve to the rainfall totals at Camden Square, London, for sets of four consecutive months, between 1870 and 1943, and found the fit to be good. In their data the frequency rises quickly to a maximum and then gradually drops down to low values.

The distribution of rainfall data for Sydney, though slightly leptokurtic, is J-shaped and therefore differs from their data. It seems to be quite commonly believed by meteorologists, however, that the log-normal curve is generally appropriate for rainfall data. In this paper we fit a log-normal curve to the Sydney rainfall data in order to see how good or bad the fit is when compared with that obtained by the use of the type III curve. To apply the maximum likelihood method it is again necessary to truncate the distribution to avoid the zero values. Maximum likelihood equations were complicated for the type III curve, but in this case they are much simpler.

Fiting of a Truncated Log-normal Curve

The log-normal probability density function is the form

$$f(x)=\frac{1}{(2\pi)^{\frac{1}{2}}\sigma(x-a)}\exp\left[-\frac{1}{2\sigma^2}\{\ln(x-a)-\mu\}^2\right]\quad\text{for } x>a\,;$$

in this case we know the origin of our distribution and so take $a=0$.

We now choose a small interval $(0, \alpha)$ and truncate the distribution at α, ignoring the actual values of x less than α, but using the fact that their total number is known. Thus, if n be the number of observations falling in $(0, \alpha)$, the rest $(N-n=m)$ of the observations will all be greater than α.

The likelihood function $\varphi(x_1, x_2, \ldots, x_m)$ in this case is given by

$$\binom{N}{n}\left[\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int_0^\alpha\frac{1}{x}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)\right\}^2 dx\right]^n\cdot\frac{1}{\sigma^m(2\pi)^{m/2}}$$

$$\times\frac{1}{\prod\limits_{i=1}^{m}x_i}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(\ln x_i-\mu)^2\right\},$$

where $x_1, x_2, \ldots, x_m \geqslant \alpha$.

Taking logarithms, we obtain

$$\begin{aligned}L&=\ln\varphi\\&=\ln\binom{N}{n}+nG(\mu,\sigma)-m\ln\sigma-\tfrac{1}{2}m\ln 2\pi-\sum_{i=1}^{m}\ln x_i-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(\ln x_i-\mu)^2,\end{aligned}$$

where

$$G(\mu,\sigma)=\ln\left[\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int_0^\alpha\frac{1}{x}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\}dx\right].$$

The maximum likelihood equations are given by

$$\frac{\partial L}{\partial \mu}=n\frac{\partial G}{\partial \mu}+\frac{1}{\sigma^2}\sum_{i=1}^{m}(\ln x_i-\mu)=0, \qquad \text{(1)}$$

$$\frac{\partial L}{\partial \sigma}=n\frac{\partial G}{\partial \sigma}-\frac{m}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{m}(\ln x_i-\mu)^2=0, \qquad \text{(2)}$$

where μ and σ are now no longer population parameters, but, for simplicity, stand for their estimates.

Now

$$\frac{\partial G}{\partial \mu}=\frac{\int_0^\alpha \frac{1}{x}\,\frac{(\ln x-\mu)}{\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\}dx}{\int_0^\alpha \frac{1}{x}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\}dx}.$$

After changing the variables in the numerator and integrating by parts, this reduces to

$$\frac{\partial G}{\partial \mu}=\frac{-\exp\left\{-\frac{1}{2\sigma^2}(\ln \alpha-\mu)^2\right\}}{\int_0^\alpha \frac{1}{x}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\}dx}. \qquad \text{(3)}$$

Further

$$\frac{\partial G}{\partial \sigma}=-\frac{1}{\sigma}+\frac{\int_0^\alpha \frac{1}{x}\,\frac{(\ln x-\mu)^3}{\sigma^3}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\}dx}{\int_0^\alpha \frac{1}{x}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\}dx},$$

which may be similarly reduced to

$$\frac{\partial G}{\partial \sigma}=-\frac{1}{\sigma}\,\frac{(\ln \alpha-\mu)\exp\left\{-\frac{1}{2\sigma^2}(\ln \alpha-\mu)^2\right\}}{\int_0^\alpha \frac{1}{x}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\}dx}. \qquad \text{(4)}$$

From (3) and (4) we see that

$$\frac{\partial G}{\partial \sigma}=\frac{1}{\sigma}(\ln \alpha-\mu)\frac{\partial G}{\partial \mu}. \qquad \text{(5)}$$

Using (5) we rewrite the likelihood equations (1) and (2) as

$$n\frac{\partial G}{\partial \mu}+\frac{1}{\sigma^2}\sum_{i=1}^{m}(\ln x_i-\mu)=0, \qquad \text{(6)}$$

$$\frac{n}{\sigma}(\ln \alpha-\mu)\frac{\partial G}{\partial \mu}-\frac{m}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{m}(\ln x_i-\mu)^2=0. \qquad \text{(7)}$$

Now multiplying (**6**) by $\frac{1}{\sigma}(\ln\alpha-\mu)$, and subtracting (**7**) from the result we find

$$\frac{1}{\sigma^3}(\log\alpha-\mu)\sum_{i=1}^{m}(\log x_i-\mu)+\frac{m}{\sigma}-\frac{1}{\sigma^3}\sum_{i=1}^{m}(\log x_i-\mu)^2=0,$$

which can be written as

$$\sigma^2=\frac{1}{m}\sum_{i=1}^{m}(\ln x_i)^2-\frac{1}{m}(\mu+\ln\alpha)\sum_{i=1}^{m}\ln x_i+\mu\ln\alpha. \quad \ldots\ldots\ldots \quad (\mathbf{8})$$

Thus equations (**6**) and (**8**) can be taken as the likelihood equations for estimating μ and σ.

To make a comparison with our fitting in the previous paper we take $\alpha=5\cdot5$. From the data we obtain

$$m=437, \qquad \sum_{i=1}^{m}\ln x_i=1373\cdot11748, \qquad \sum_{i=1}^{m}(\ln x_i)^2=4717\cdot25359.$$

Substituting these values in (**6**) and (**8**), we find that (**8**) reduced to

$$\sigma^2=5\cdot43807-(1\cdot43739)\mu, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{9})$$

and (**6**) reduces to

$$n\frac{\partial G}{\partial\mu}+\frac{1}{\sigma^2}(1373\cdot11748-437\mu)=0. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{10})$$

Solving (**9**) and (**10**) we find $\mu=-0\cdot17$ and $\sigma=2\cdot38$ correct to two places of decimals.

The calculation of the expected frequencies for testing the goodness of fit is shown in Table 1. Since the χ^2 test is used for testing goodness of fit, the observations are grouped into classes so that the expected frequency in any class is not less than 5. To facilitate comparison we have here grouped the observations into the same classes as we did in our previous paper for fitting a truncated type III distribution. We have used the Tables of Probability Functions, Vol. I (National Bureau of Standards 1941) to calculate the expected frequencies.

The results are shown in the table. The first column gives the class interval, the column headed f_0 gives the corresponding observed class frequencies, and the column headed $f_E^{(1)}$ gives the expected frequencies based on log-normal distribution. Thus for 16 degrees of freedom the total χ^2 is found to 44·0, which shows that the fit is a very poor one. On the other hand the column headed $f_E^{(2)}$ gives the frequencies expected on the basis of a truncated type III curve. The value of χ^2 is here found to be 7·8, showing that the fit is an extremely good one.

Finney (1941) showed that, if the variable x is such that $\log x$ is normally distributed with mean ξ and variance σ^2, then the x population has the mean $\exp(\xi+\delta\sigma^2)$ and variance $\exp(2\xi+\sigma^2)(\exp\sigma^2-1)$. Accordingly the estimates of the mean and standard error of the daily rainfall distribution based on log-normal curve are given by

$$\text{mean} = 14\cdot4 \text{ and standard deviation} = 246\cdot2.$$

Based on type III curve these estimates are

$$\text{mean} = 8\cdot1 \text{ and standard deviation} = 24\cdot9.$$

But calculating directly from the sample we get

$$\text{sample mean} = 8\cdot4 \text{ and sample standard deviation} = 27\cdot9.$$

Thus the mean and standard deviation calculated directly from the sample agree well with those estimated on the basis of the type III curve, but deviate considerably from those estimated on the basis of the log-normal curve. As is evident from the table, this is due to the fact that the frequency towards the upper tail of the log-normal curve is very much higher than that observed. This has resulted in increasing the mean, and, more especially, the standard deviation.

Table 1

FREQUENCIES AS PREDICTED BY THE FITTED CURVES

Class Interval	f_0	$f_E^{(1)}$	$f_E^{(2)}$	Class Interval	f_0	$f_E^{(1)}$	$f_E^{(2)}$
0– 5	1631	1621·6	1638·5	51– 60.. ..	18	13·3	19·6
6–10	115	146·4	106·0	61– 70.. ..	13	10·2	14·7
11–15	67	70·1	62·0	71– 80.. ..	13	7·7	11·6
16–20	42	42·7	43·6	81– 90.. ..	8	6·3	8·9
21–25	27	29·1	32·2	91–100.. ..	8	5·1	7·2
26–30	26	21·1	26·0	101–125.. ..	16	9·4	12·2
31–35	19	16·1	20·7	126–150.. ..	7	6·4	7·2
36–40	14	12·7	17·2	151–225.. ..	9	10·9	9·5
41–45	12	10·5	14·3	226 or more ..	5	19·8	4·4
46–50	18	8·6	12·2				

On the basis of these results, it would seem that, at any rate when a certain proportion of days have zero rainfall, the better curve to use in fitting rainfall data is the type III curve, not the log-normal, as has been previously assumed by several writers in meteorology.

The author thanks Professor P. A. Moran for suggesting the problem to him.

References

Blackhouse, T. W. (1891).—*Quart. J. R. Met. Soc.* **17**: 286.

Bowen, E. G. (1953).—*Aust. J. Phys.* **6**: 490–7.

Brooks, C. E. P., and Carruthers, N. (1953).—" Handbook of Statistical Methods in Meteorology." p. 102. (H.M. Stationery Office: London.)

Das, S. C. (1955).—*Aust. J. Phys.* **8**: 298–304.

Finney, D. J. (1941).—*J. R. Stat. Soc. Suppl.* **7**: 155–61.

National Bureau of Standards (1941).—" Tables of Probability Functions." Vol. 1.

COMPOSITION OF THE ^{7}Li(p,γ) RADIATION BELOW THE 441 KEV RESONANCE*

By J. G. Campbell†

The only published determinations of the energy of the γ-radiation emitted when lithium is bombarded by protons with energy less than 400 keV have been made by absorption measurements of secondary electrons in aluminium. Gentner (1937) found that the radiation at 90° from a thick lithium target at 300 keV had the same half-value thickness as that at 550 keV, corresponding to about 17 MeV. Tangen (1946), however, reported that the radiation emitted at 0° at both 300 and 400 keV had energy $14 \cdot 5 \pm 0 \cdot 5$ MeV.

The later work of Walker and McDaniel (1948) has shown that the radiation from this reaction consists essentially of two components, at 17·6 and at 14·8 MeV. This note reports a scintillation counter investigation of the ratio of these components at low proton energies. Three other components of lower energy, with combined intensity at resonance amounting to 5 per cent. of the total, have also been identified, by Inall and Boyle (1953), but on the assumption that they remain proportionately small off resonance their effect is neglected here.

A sodium iodide crystal $1\frac{1}{2}$ in. long and $1\frac{1}{2}$ in. in diameter was placed with its centre 3 in. from the lithium target, its axis being on a line from the target at 90° to the direction of the proton beam. A 6260 photomultiplier was coupled optically to it and fed pulses through amplifiers to both a 1009 discriminator-scaler and a single-channel analyser in parallel. The discriminator and the analyser were intercalibrated to ensure that their readings were strictly comparable on the same scale.

The differential pulse-height distribution for thick-target lithium resonance radiation is shown in Figure 1. The continuous spread of the distribution is due to bremsstrahlung escape from the crystal (Campbell and Boyle 1954). The extrapolation to zero pulse height is taken as a straight horizontal line, to eliminate the usual spurious low energy rise.

The assumption is now made that the two components of the radiation give rise to individual pulse-height distributions of the same shape, the maximum pulse height in each distribution being proportional to the photon energy. This assumption is justified by the fact that the greater bremsstrahlung spreading of the 17·6 MeV component is compensated by the natural line breadth of 2 MeV over which the 14·8 MeV component is spread.

* Manuscript received November 7, 1955.

† Physics Department, University of Melbourne; present address: Australian Paper Manufacturers Ltd., Melbourne.

In order to resolve the resultant distribution into the two individual contributions it is necessary to know the value of I_1/I_2, the ratio of intensities of the 17·6 to the 14·8 MeV component. The best published value of this ratio at resonance (in which case both the angular distributions are isotropic) is 1·7 ±0·2, which is given by Stearns and McDaniel (1951) and includes the earlier work of Walker and McDaniel (1948). The result of a graphical analysis using this value is shown by the broad band in Figure 1, the depth of the hatching indicating the limits set by the uncertainty in I_1/I_2.

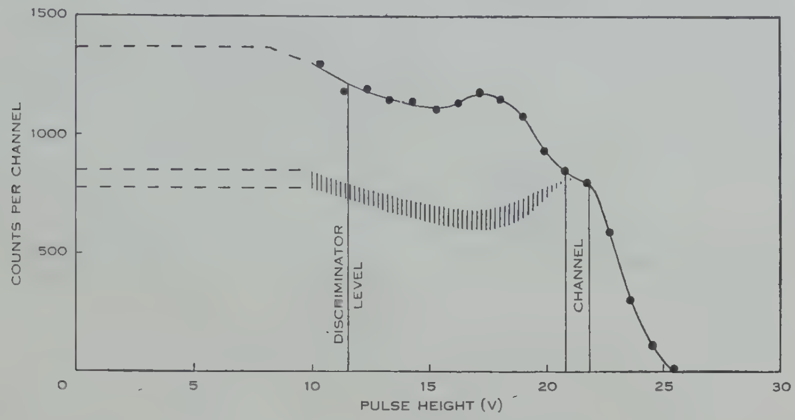


Fig. 1.—Pulse-height distribution at resonance.

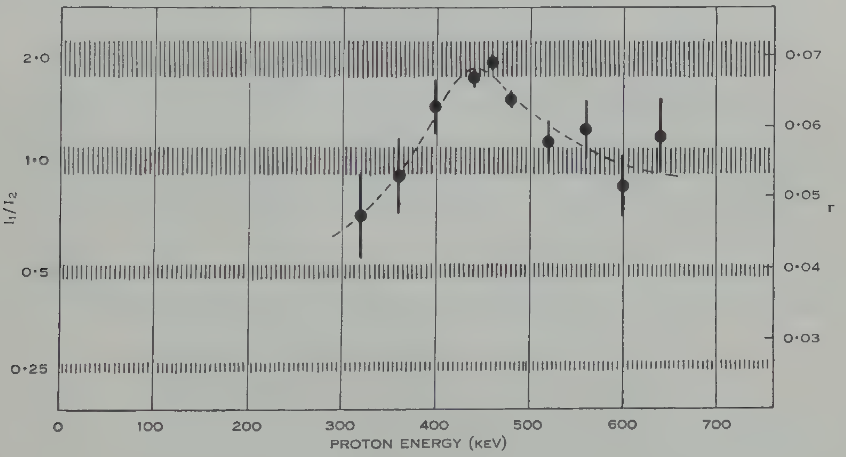


Fig. 2.—Results for thin target.

The discriminator level was set at 11·5 V, where it registered counts due to both γ-components, and the 1-V channel of the analyser was set with its centre at 21·3 V, where it responded almost exclusively to the 17·6 MeV component. From the ratio of the two counts, the areas in Figure 1 to which they correspond, and the known efficiencies of the crystal at the two photon energies, the ratio I_1/I_2 can be determined at other proton energies.

A thin (~10 keV) target of lithium metal was first bombarded by the proton beam, and the ratio r (analyser counts divided by discriminator counts) measured at several beam energies. The results are shown in Figure 2. The ordinate in this figure is a linear scale in r, as shown along the right-hand edge. The left-hand edge is calibrated in I_1/I_2, and the vertical hatching, corresponding to the hatching in Figure 1, is due to the uncertainty in the value of this at resonance. The behaviour of the experimental points, rising to a maximum at the 441 keV resonance and then dropping back to about $I_1/I_2=1$, is in agreement with that observed at 90° between 400 and 600 keV by Devons and Hine (1949). This run therefore acts as a check on the validity of the method.

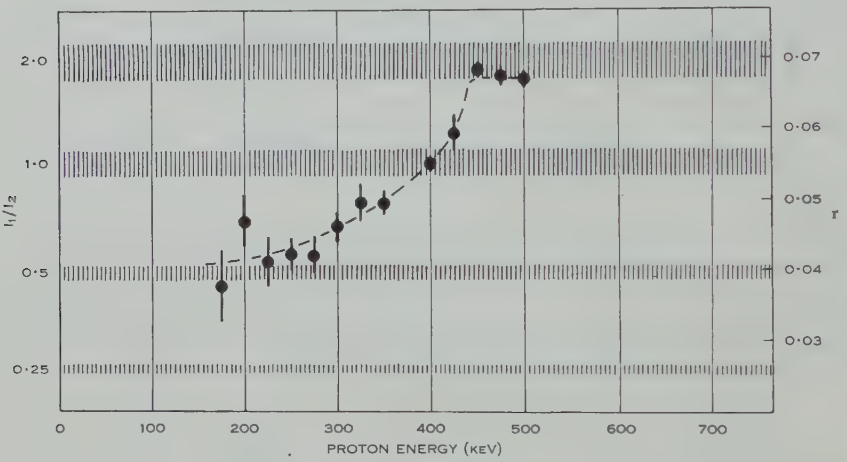


Fig. 3.—Results for thick target.

The results for a thick lithium target at low proton energies are now plotted in Figure 3. Owing to the steeply rising nature of the reaction cross section up to the resonance (Tangen 1946) the thin-target spectrum can be taken to be substantially similar to this.

It is evident from Figure 3 that, as the proton energy is reduced below resonance, the proportion of 17·6 MeV radiation decreases sharply, until at 200 keV it is only about half as intense as the 14·8 MeV radiation.

References

Campbell, J. G., and Boyle, A. J. F. (1954).—*Aust. J. Phys.* **7**: 284.
Devons, S., and Hine, M. G. N. (1949).—*Proc. Roy. Soc.* A **199**: 56.
Gentner, W. (1937).—*Z. Phys.* **107**: 354.
Inall, E. K., and Boyle, A. J. F. (1953).—*Phil. Mag.* **44**: 1081.
Stearns, M. B., and McDaniel, B. D. (1951).—*Phys. Rev.* **82**: 450.
Tangen, R. (1946).—K. Norske Vidensk. Selsk. Skr. No. 1.
Walker, R. L., and McDaniel, B. D. (1948).—*Phys. Rev.* **74**: 315.

IONOSPHERIC RECORDERS AND SPORADIC E[*]

By J. A. Thomas,[†] A. C. Svenson,[†] and H. E. Brown[‡]

It has been many times stated (Rawer 1949; Booker 1950; Booker and Gordon 1950; McNicol and Gipps 1951) that the frequency characteristics of the sporadic E region are rapidly varying functions of the overall sensitivity of the recording $P'f$ equipment. A very simple test of this statement has been made, and it has been found to be incorrect at Brisbane.

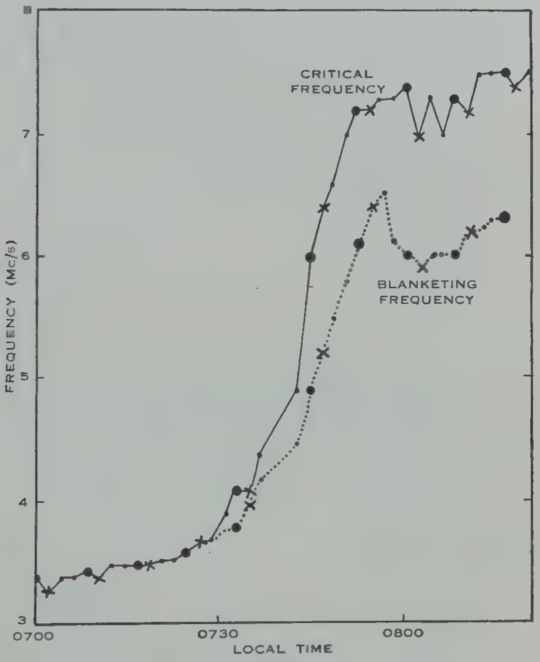


Fig. 1.—Variation of fE_s and f_bE_s for sequential type E_s for different recorder sensitivities (November 16, 1955).
● 0 dB × −24 dB · −12 dB

The tests were made by taking repetitive sequences of records (on an auxiliary display) with different receiver gains. Each two minutes the receiver gain was automatically changed by a known factor; any resultant changes in fE_s or f_bE_s should thus show clearly when the scaled values are plotted over a period of time. Two different sensitivity sequences have been used: (a) −12 dB,

---

[*] Manuscript received December 15, 1955.

[†] Physics Department, University of Queensland, Brisbane.

[‡] Ionospheric Prediction Service, c/o Physics Department, University of Queensland, Brisbane.

0 dB, −24 dB, −12 dB; and (*b*) −12 dB, +6 dB, −24 dB, −6 dB. (A reflection coefficient of about 0·003 is necessary to produce a record at the 0 dB level.) A typical sequence is shown in Plate 1; the camera speed is one-third of its normal value.

The scaled values of fE_s and f_bE_s for typical cases of the two types of E_s which occur at Brisbane are shown plotted in Figures 1 and 2. In neither case is there any evidence of dependence of fE_s on the receiver sensitivity (and hence on overall equipment sensitivity). This result is to be expected for the sequential type of E_s (McNicol and Gipps 1951), but is now established for the constant-height type of E_s *as it occurs at Brisbane.* Provided E_s is recorded at all it will be recorded with a critical frequency which is independent of overall sensitivity up to at least 6 dB above the normal recording sensitivity. Blanketing frequency is also independent of sensitivity—such variations as do occur on change of sensitivity are no larger than the random changes constantly occurring on fixed sensitivity.

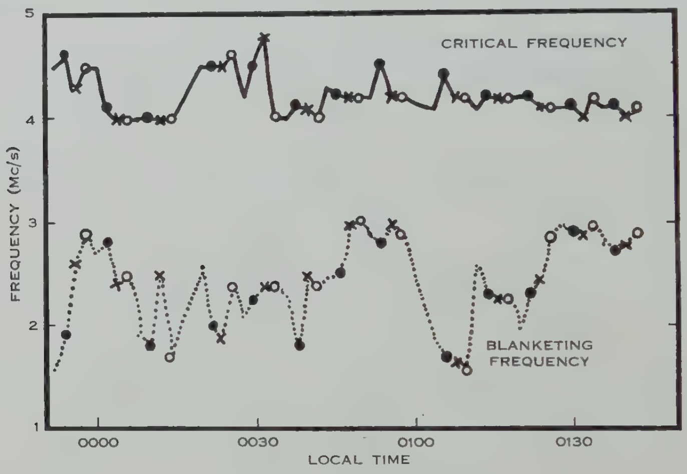


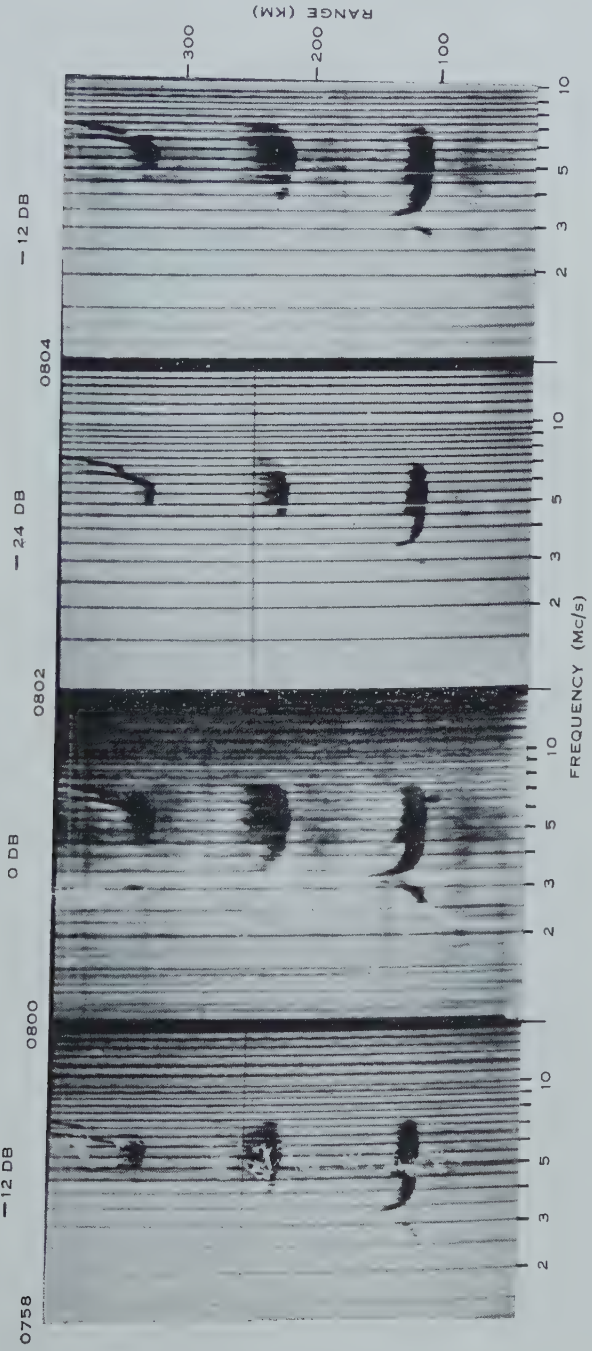
Fig. 2.—Variation of fE_s and f_bE_s for constant height type E_s for different recorder sensitivities (November 16, 1955).
● +6 dB ○ −6 dB · −12 dB × −24 dB

These results may mean that more reliance can be placed on the fE_s and f_bE_s values for stations in mid latitudes than has been thought possible in the past; the marked variations between stations may be due to the varying nature of the sporadic E itself rather than to variations in overall recording sensitivity between the stations.

The authors wish to thank Mr. G. Cairns for his assistance in taking the measurements.

This work forms part of the research programme of the Radio Research Board of C.S.I.R.O. and of the Ionospheric Prediction Service of the Commonwealth Observatory, and is published by permission of these bodies.

IONOSPHERIC RECORDERS AND SPORADIC E



A typical sequence of auxiliary $P'f$ records with varying sensitivity (November 16, 1955).

References

Booker, H. G. (1950).—Proc. Pennsylvania Conf. Ionospheric Physics *A*, p. 171.
Booker, H. G., and Gordon, W. E. (1950).—*Proc. Inst. Radio Engrs., N.Y.* **38** : 401.
McNicol, R. W. E., and Gipps, G. de V. (1951).—*J. Geophys. Res.* **56** : 17.
Rawer, K. (1949).—*Nature* **163** : 528.

PROCESSES CONTROLLING IONIZATION DISTRIBUTION IN THE F_2 REGION OF THE IONOSPHERE*

By D. F. Martyn†

It has been known for more than 20 years that the morphology of the principal (F_2) region of the ionosphere is complex, and inexplicable in terms only of solar ionizing radiation and recombination. The hypothesis was advanced (Martyn 1947) that movement, and especially vertical movement, of F_2 ionization was the main reason why the morphology of this region differed so profoundly from that of the lower regions E and F_1, which conform closely with Chapman's well-known theory. A principal cause of movement of F_2 ionization was held to be the drift occasioned by the electrodynamic forces and winds associated with the ionospheric electric current systems whose existence is manifested by the daily magnetic variations at the ground. There is now considerable evidence, both theoretical and observational, to show that this view may be essentially correct.

During the past few years rockets have been used, especially in the United States, in attempts to measure *in situ* the physical properties of the high atmosphere. In particular, a small number of measurements of atmospheric density has been made at heights approaching 200 km. These suggest that the density of the atmosphere in the F_2 region may be some 30 times less than had hitherto been deduced (Martyn and Pulley 1936) from radio experiments. If the air density at these levels were as low as rocket measurements suggest, it would be likely that diffusion of F_2 ionization (in electrically neutral volume) under the forces of gravity and of its own (partial) pressure gradient would notably affect the shape and height of the region. Accordingly, an investigation has been made of the possible effects of diffusion processes at such levels.

The results of this investigation, as outlined below, lead to the following conclusions.

(*a*) The rocket measurements of atmospheric densities at heights above 150 km appear to be subject to a gross systematic error, and are much lower than can be reconciled with other types of observation.

---

* Manuscript received January 13, 1956.

† Radio Research Laboratory, C.S.I.R.O., Camden, N.S.W.

(*b*) A Chapman (C_α) region, defined as one in which the ion pairs are produced by absorption of monochromatic radiation in an isothermal atmosphere and in which they decay by recombination, has a characteristic form given by

$$N=N_m \exp \tfrac{1}{2}\{1-z-e^{-z}\}, \quad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

where N is the number density of the ion pairs and z is height, measured from the level of maximum ionization, in units of H,* the " scale height " of the atmosphere. It is found that a region of any initial shape will in time assume by diffusion the form of a C_α region ; the time required for this to happen at F_2 heights (*c*. 350 km) is a few hours, if the air density there is as great as was formerly believed, but only a few minutes if the rocket-measured densities are correct.

(*c*) It is found that, if the decay of electron density occurs according to the " attachment " formula

$$\partial N/\partial t=-\beta_0 e^{-z}N,$$

then in moderate to high magnetic latitudes a stable and stationary C_α region will tend to form at one particular height, which is 360 km if pre-rocket densities are accepted ; this is observed to happen during the night hours in all parts of the world save the immediate vicinity of the magnetic equators.

(*d*) It is found that the formation of a C_α region by diffusion a few hours after sunset accounts for the well-known fact that at such times the under surface of the F_2 region assumes a parabolic height distribution.

(*e*) Finally, the present investigation confirms an earlier suggestion (Martyn 1954) that the well-known geomagnetic equatorial anomaly is due to the combination of large vertical electrodynamic drift velocities and inhibition of vertical diffusion by the nearly horizontal magnetic field. Thus by day the F_2 region near the magnetic equators is drawn out by upward drift until it becomes a high thick region of abnormally small peak ionization, while the shape of the under surface has no resemblance to parabolic form. During the early part of the night the peak density is sustained because the rate of disappearance of ionization is small at the great heights in question. Eventually the peak is lowered, by electrodynamic drift, to heights less than 300 km, where the relaxation time is of the order of 1 hr, and decay is rapid.

Speed of Diffusion of Ionization under Gravity and Partial Pressure Gradient

It is a safe assumption that in the ionosphere at a given point the number densities of the positive and of the negative particles N_i, N_e are almost exactly equal ; even a small departure from equality would give rise to very strong electric fields which would immediately restore the balance. It is also safe to assume, for the low pressures of the F_2 region, that transport due to mechanical forces can occur only along the lines of geomagnetic force. If a particle of mass m, making an average of ν collisions per second with the surrounding gas molecules, is influenced by a force F it will drift through the gas with the mean velocity

---

* In this paper g, the acceleration due to gravity, is measured in similar units.

$$v = F/m\nu. \qquad (2)$$

It is clear that the positive ions must drift with the same speed as the electrons, the more mobile particles (the electrons) being tied to the ions by powerful electrostatic fields. Thus, since $m_i\nu_i \gg m_e\nu_e$, we need only determine the speed of the ions in order to find the speed of the ionization transport.

The vertical gravity force upon an ion pair is $-(m_i+m_e)g$, or, to close approximation $-m_i g$. The vertical force upon unit volume of ionization due to its own pressure gradient is $-\partial(p_i+p_e)/\partial h$, where $p_i=p_e=NkT$, T being the mean absolute temperature of the ionization and k Boltzmann's constant. Thus this force is $-2kT\partial N/N\partial h$ per ion pair. It is clear that it will be almost wholly transferred to the less mobile particle (the ion) by electrostatic interaction. The transport speed due to gravity and pressure gradient is therefore

$$\begin{aligned} v &= -\frac{m_i g}{m_i\nu_i} - \frac{2kT}{Nm_i\nu_i} \cdot \frac{\partial N}{\partial h} \\ &= -\frac{g}{\nu_i} - \frac{2gH}{\nu_i N} \cdot \frac{\partial N}{\partial h} \\ &= -\frac{g}{\nu_i}\left(1 + \frac{2}{N}\,\frac{\partial N}{\partial z}\right). \qquad (3) \end{aligned}$$

Since ionization transport can occur only along geomagnetic lines of force this equation as it stands will apply only at the magnetic poles. In regions where the magnetic inclination is ψ the effective force is $F \sin \psi$, and the *vertical* velocity is reduced by the factor $\sin^2 \psi$.

The above equation has previously been derived by other workers by more complex reasoning. The above simple derivation may help to clarify the physical processes involved.

Formation of Stable C_α *Regions by Diffusion*

The continuity equation for a changing vertical distribution of ionization in the F_2 region in the absence of sunlight and decay is

$$\frac{\partial N}{\partial t} = -\operatorname{div}(Nv). \qquad (4)$$

A region able to move vertically (or to remain stationary) without distortion must satisfy the condition

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -N\frac{\partial v}{\partial z} - v\frac{\partial N}{\partial z} + v\frac{\partial N}{\partial z} = -N\frac{\partial v}{\partial z} = 0$$

at all heights.

In other words v must have no height gradient. Thus, from (3), writing φ for $(1/N)\partial N/\partial z$, we must have

$$\frac{\partial \varphi}{\partial z} + \varphi = -\tfrac{1}{2}. \qquad (5)$$

The solution of (5) is equation (1) above, which is the ionization distribution of a C_α region. Also, from (3) we find

$$v = -g \sin^2 \psi / \nu_0,$$

where ν_0 is the ion collision frequency at the height $z=0$, where $N = N_m$.

Clearly, the shape of a C_α region remains unaffected by diffusion, a conclusion already reached (Martyn 1954; Hirono 1955) by other methods. It does not necessarily follow, however, that such a region is *stable*, or that other distributions of ionization will assume the C_α form.

The stability of a C_α region has been investigated by a perturbation method, the details of which will not be given here; it is found that if ΔN be the perturbation, then $(1/\Delta N)\partial(\Delta N)/\partial t$ is negative in sign, thus demonstrating that the region is stable for small disturbances.

To test the second aspect a number of different initial distributions, such as $N = N_m \exp(-bz^2)$, were assumed. In all cases it was found that the ionization became redistributed in such a way as to form a C_α region. There seems little reasonable doubt that any arbitrary ionization distribution having an ionization peak will eventually, in a time of order $\nu_0/g \sin^2 \psi$ assume the form of a C_α region.

Formation of a Stable C$_\alpha$ *Region at a Particular Height, by Combined Action of Diffusion and Attachment Gradient*

In order to simulate actual conditions in the F_2 region at night, it is necessary to include in equation (4) a term which will take account of the decay* of ionization. Thus we write

$$\frac{\partial N}{\partial t} = -\text{div}\,(Nv) - \beta_0 e^{-z}.$$

It is important to enquire whether it is possible for a decaying C_α region to form at any particular level in the ionosphere.

The condition for this to occur would be

$$\frac{1}{N}\frac{\partial N}{\partial t} = -B,$$

where B is a constant, independent of height and time. It is readily found that a C_α region can exist in a stable, stationary (though decaying), state, with its peak at a height given by $2\beta\nu = g \sin^2 \psi$; that is to say, (taking ν as $7{\cdot}6$ at 300 km and H as 50 km) at a height of 359 km.

It is a well-known fact that the F_2 region rises rapidly after sunset, and settles for some hours with its peak at approximately this height, in all parts of the world save the immediate vicinity of the magnetic equator. The processes just described would be inhibited by the horizontal magnetic field near this equator. Thus it would appear that the observed behaviour of the F_2 region after sunset is consistent with the view that it is substantially controlled by the process just outlined.

---

* Whether F_2 ionization decays according to a recombination or attachment law has been debated for many years. Ratcliffe *et al.* (1956) have produced strong evidence that an attachment law is effective, and that, between heights of 250 and 350 km, $\beta = 10^{-4} \exp-[\{h\ (\text{km}) - 300\}/50]$. By different methods, still unpublished, the present author has verified both these conclusions, and finds $\beta = 10^{-4} e^{-z}$.

Interpretation of Equatorial Anomalies, and of Parabolic Shape of F_2 *Region*

It is strikingly noticeable that the daily variation of N_m and z_m within 10–15° of the magnetic equators is very different from the corresponding variations at higher latitudes. While it is true that vertical electrodynamic drift is larger in low latitudes, it would be expected theoretically to decrease to a shallow minimum at the equator. Observations show no sign of such a minimum.

The deductions made above appear to confirm the suggestion (Martyn 1954) that *absence* of vertical diffusion at the equators is responsible for the very great height variations, and the abnormal N_m variations, there observed. The above reasoning leads to the conclusion that diffusion would prevent the F_2 region from being elevated above about 400 km in moderate to high latitudes, and even this height should seldom be attained. Moreover, the fact that diffusion leads to grouping of ionization in C_α form would tend to prevent the F_2 region from being abnormally attenuated or spread out over a large height range, with consequent reduction in N_m. No such considerations hold near the equators, where there is nothing to prevent the formation of " thick " F_2 regions of abnormally reduced N_m, and with peaks as high as 600 or more km.

Equation (**1**) may be approximated, for values of $|z|$ not much greater than unity, to

$$\begin{aligned} N &\approx N_m\{1+\tfrac{1}{2}-\tfrac{1}{2}z-\tfrac{1}{2}(1-z+\tfrac{1}{2}z^2)\} \\ &\approx N_m(1-\tfrac{1}{4}z^2). \end{aligned}$$

This is the parabolic form which the under surface of the region assumes an hour or two after sunset. It would appear that this well-known phenomenon is simply another aspect of the tendency of the region to assume C_α form when the ionizing agency is removed. This view is confirmed by the observed fact that the region does not assume this form near the magnetic equators.

The Density of the Atmosphere in the F_2 *Region*

The simple interpretation of all the above well-known F_2 phenomena assumes that the atmospheric densities derived from radio data are substantially correct, the air particle density being close to 10^{10} per c.c. at the 300 km level. On the other hand, if recent rocket-measured densities were accepted, the F_2 region would, at moderate to high latitude, move down after sunset to well below 300 km. This happens nowhere, and suggests that there must be present a systematic source of error in the devices used for density measurement by rocket at great heights.

References

Hirono, M. (1955).—*Rep. Ionosphere Res. Japan* **9**: 95–104.

Martyn, D. F. (1947).—*Proc. Roy. Soc.* A **189**: 241–60.

Martyn, D. F. (1954).—" The Physics of the Ionosphere." pp. 254–9. (Phys. Soc.: London.)

Martyn, D. F., and Pulley, O. O. (1936).—*Proc. Roy. Soc.* A **154**: 455–86.

Ratcliffe, J. A., Schmerling, E. R., Setty, G. S. G. K., and Thomas, J. O. (1956).—*Phil. Trans.* A **248**: 621.

CONDUCTION OF HEAT IN AN INFINITE REGION BOUNDED INTERNALLY BY A CIRCULAR CYLINDER OF A PERFECT CONDUCTOR

By J. C. JAEGER*

[*Manuscript received November* 24, 1955]

Summary

Numerical information is given for radial flow of heat in an infinite region bounded internally by a circular cylinder of radius a containing a perfect conductor, there being contact resistance $1/H$ per unit area across the cylinder. Problems considered are: (i) the perfect conductor and the surrounding region initially at different temperatures; (ii) heat supply to the perfect conductor. Results are expressed in terms of the dimensionless parameters $\tau=\varkappa t/a^2$, $h=K/aH$, $\alpha=2\pi a^2\rho c/S$, where K, ρ, c, $\varkappa$ are the thermal conductivity, density, specific heat, and diffusivity in the region outside the cylinder, S is the thermal capacity of the perfect conductor per unit length of the cylinder, and t is the time. Tables and graphs of the temperature of the perfect conductor are given for $0\cdot 2<\tau<20$; $h=0, 0\cdot 5, 1, 2, 3, 4, 5, 7, 10, 20$; and $\alpha=0\cdot 5, 1, 1\cdot 5, 2$. The temperature outside the cylinder, a problem involving fluid motion within the cylinder, and the heating of a buried cable carrying electric current are also discussed.

I. INTRODUCTION

Radial flow of heat in the infinite region surrounding an infinite circular cylinder which contains a perfect conductor or well-stirred fluid has been discussed at some length as an example of an important boundary condition in conduction of heat (Jaeger 1940). Problems of this type are of great importance since they provide good approximations to many important practical problems such as the heating of a buried cable, while the transient heating of a solid by a wire carrying current has frequently been proposed as a method for measuring thermal conductivity (Fischer 1939; van der Held, Hardebol, and Kalshoven 1953). A better representation of experimental conditions is frequently obtained by adding the additional complication of a contact resistance between the perfect conductor and the surrounding solid.

The solutions of all the problems considered below may readily be obtained by the method of the Laplace transformation (cf. e.g. Carslaw and Jaeger 1947). Unfortunately, rather complicated integrals are involved which have to be evaluated numerically. Approximations useful for large or small values of the time may also always be found by the usual methods, and, because of the absence of accurate numerical information, there has been a tendency to use these in the reduction of experimental results. Physically, they frequently correspond to replacing the problem by the simpler one of a line source in an infinite medium.

---

* Australian National University, Canberra, A.C.T.

In some cases this is satisfactory, but the parameters involved (in particular the times concerned) are often such as to make them quite inapplicable.

Two particular problems occur in many contexts. The first of these is that in which the perfect conductor and the surrounding material are initially at different temperatures. This has been discussed (for the case of no contact resistance) using an approximate method by Whitehead (1944), in connexion with the dissipation of heat from a suddenly heated electrical cable. Jaeger (1940), Tranter (1947), and Bullard (1954) have given the exact solution, and the latter has given some numerical values ; he also points out the possibility of using these results for the measurement of thermal conductivity. This problem, including the case of contact resistance, is discussed in Sections III and IV.

In the second problem, the perfect conductor and its surroundings are initially at zero temperature and the perfect conductor is subsequently heated at a constant rate. This has been discussed approximately in connexion with the heating of electric cables by Whitehead and Hutchings (1938). van der Held, Hardebol, and Kalshoven (1953) use the exact solution with no contact resistance for the measurement of thermal conductivity. Blackwell (1954) has given the solution for the case in which there is contact resistance, and uses the first two terms of its asymptotic expansion in the determination of thermal conductivity. Numerical information about this problem is given in Section V.

The application of these results to the problem of the heating of a buried cable, and the solution of a more general problem of this type, are discussed in Section VII.

The case in which the cylinder contains well-stirred fluid and heat is supplied to it by exchange of fluid has been studied by Jaeger (1940). Some numerical results are given in Section VI.

In most cases numerical results are given only for the surface temperature of the solid. Values for the temperature in the interior of the solid may be obtained from these results by the use of Duhamel's theorem (Carslaw and Jaeger 1947, Section 10) and tables given by Jaeger (1956).

The calculations given here were made in connexion with the measurement of thermal conductivity of rocks from boreholes. This application is discussed in a companion paper (Beck, Jaeger, and Newstead 1956).

II. NOTATION

In all cases the region outside the circular cylinder $r=a$ will be supposed to contain solid of density ρ, specific heat c, thermal conductivity K, and diffusivity $\kappa=K/\rho c$; v will be written for the temperature in this material at distance r from the axis at time t, the material being always supposed to be initially at zero temperature ; v_s will be written for the surface temperature of this region, that is, the value of v when $r \to a+0$.

The cylinder $r=a$ is supposed to contain a perfect conductor (or well-stirred fluid), its heat capacity per unit length of the cylinder being S. V will be written for its temperature. In some problems heat will be supplied to the perfect conductor at the rate Q per unit time per unit length.

At the surface $r=a$ it will be assumed that there is a thermal contact resistance $1/H$ per unit area between the perfect conductor and the surrounding solid, so that the flux of heat outwards across this surface is

$$H(V-v_s). \qquad \text{(1)}$$

The case $H\to\infty$ gives perfect thermal contact at the surface.

The theoretical formulae may be expressed in terms of three dimensionless parameters, namely,

$$\tau=\varkappa t/a^2, \qquad \text{(2)}$$

$$\alpha=2\pi a^2\rho c/S, \qquad \text{(3)}$$

$$h=K/aH. \qquad \text{(4)}$$

The numerical values of these likely to arise in practice will now be discussed. When units are involved they will be c.g.s., calorie, and °C. Only poor conductors need be considered, since if the region $r>a$ contains a good conductor the assumption that $r<a$ contains a perfect conductor will cease to be a reasonable approximation to any likely physical situation.

For an experiment of 10 min duration in a hole of radius $a=2$ cm in material of diffusivity $\varkappa=0\cdot01$, (**2**) gives $\tau=1\cdot5$ so that in such experiments the range $0\cdot2$–10 of τ may be regarded as important. If, however, the radius is reduced to 2 mm, the values of τ become large and simple approximate formulae are available.

The parameter α is twice the ratio of the heat capacity of a cylinder of the solid of radius a and unit length to that of unit length of the cylinder of perfect conductor. If the perfect conductor is well-stirred water and it is surrounded by rock, α is about 1; for solid brass in rock, α is about $1\cdot5$; for the composite probe used by Bullard (1954), $\alpha=2$; for brass in a lighter material such as wood, α falls to $0\cdot3$. If a tube of perfect conductor is used instead of a solid cylinder higher values of α occur. It may be noted here that if $\alpha=2$ the asymptotic expansion for large values of the time takes a slightly simpler form so that there is some reason for approximating to this case when designing apparatus.

The likely values of h may be estimated by considering the case of an air gap 1 mm thick and of conductivity $0\cdot000053$ in a hole of radius $a=2$ cm in rock of conductivity $K=0\cdot0053$. This gives $h=5$. Thus, for a slightly irregular hole in dry rock, values of h at least as large as this must be allowed for. On the other hand, if the hole is filled with well-stirred fluid or there is really good contact, as in the case of Bullard's experiments, it is reasonable to neglect contact resistance and to take $h=0$.

III. THE CASE OF DIFFERENT INITIAL TEMPERATURES

Suppose the region $r<a$ to be initially at temperature V_0 and the region $r>a$ at zero temperature, no heat being supplied to the system.

Then the temperature V for $r<a$ at time t is given by

$$V/V_0=F(h,\alpha,\tau), \qquad \text{(5)}$$

where

$$F(h,\,\alpha,\,\tau)=\frac{4\alpha}{\pi^2}\int_0^\infty \frac{\exp\,(-\tau u^2)\mathrm{d}u}{u\Delta(u)}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

$$\Delta(u)=[u\mathrm{J}_0(u)-(\alpha-hu^2)\mathrm{J}_1(u)]^2+[u\mathrm{Y}_0(u)-(\alpha-hu^2)\mathrm{Y}_1(u)]^2, \quad \ldots (7)$$

τ, α, h are defined in (2)–(4), and $\mathrm{J}_n(u)$ and $\mathrm{Y}_n(u)$ are the Bessel functions of order n of the first and second kinds.

This result is easily obtained by the Laplace transformation method as in Jaeger (1940). For the important case of perfect thermal contact, $h=0$, it reduces to one given therein and also by Bullard* (1954). Values for the case $h=0$ are given in Table 1 for various values of τ and α.

TABLE 1

VALUES OF $F(0,\,\alpha,\,\tau)$

α τ	0·5	1·0	1·5	2·0	4·0	6·0	8·0
0·2	0·757	0·595	0·482	0·401	0·229	0·156	0·118
0·3	0·707	0·529	0·414	0·335	0·182	0·122	0·091
0·4	0·668	0·480	0·365	0·290	0·152	0·101	0·075
0·5	0·634	0·440	0·328	0·257	0·132	0·087	0·064
0·6	0·604	0·408	0·298	0·231	0·116	0·076	0·056
0·7	0·578	0·380	0·274	0·210	0·104	0·068	0·050
0·8	0·555	0·357	0·253	0·193	0·095	0·062	0·046
0·9	0·534	0·336	0·236	0·179	0·087	0·057	0·042
1·0	0·514	0·317	0·221	0·166	0·080	0·052	0·039
2·0	0·381	0·206	0·135	0·099	0·047	0·030	0·022
3·0	0·302	0·152	0·098	0·071	0·033	0·022	0·016
4·0	0·250	0·120	0·076	0·055	0·026	0·017	0·012
5·0	0·212	0·099	0·062	0·045	0·021	0·014	0·010
6·0	0·184	0·084	0·053	0·038	0·018	0·012	0·009
7·0	0·162	0·073	0·046	0·033	0·016	0·010	0·008
8·0	0·144	0·064	0·040	0·029	0·014	0·009	0·007
9·0	0·129	0·057	0·036	0·026	0·012	0·008	0·006
10·0	0·117	0·052	0·033	0·024	0·011	0·007	0·006
15·0	0·079	0·035	0·022	0·016	0·008	0·005	0·004
20·0	0·059	0·026	0·017	0·012	0·006	0·004	0·003

For the case in which there is contact resistance, three parameters are involved and values for the most important cases $\alpha=1$ and $\alpha=2$ are shown in Figures 1 and 2. These values are sufficient for many practical purposes; for other values of α curves can be constructed by interpolation. The plot of log F against log τ was introduced by Bullard and is most convenient, both for display and practical applications.

---

* The notation is chosen to correspond with Bullard's; $F(0,\,\alpha,\,\tau)$ is his $F(\alpha,\,\tau)$. He gives values of this function to 4 D for $\alpha=2,\,4,\,6,\,8$, and $\tau=0{\cdot}2(0{\cdot}2)3(1)10$.

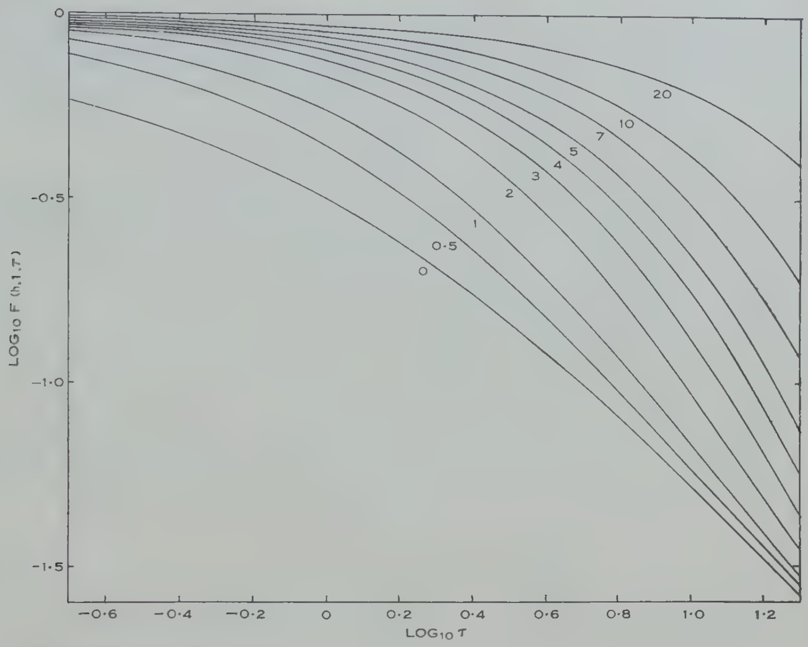


Fig. 1.—The function $F(h, 1, \tau)$. Numbers on the curves are the values of h.

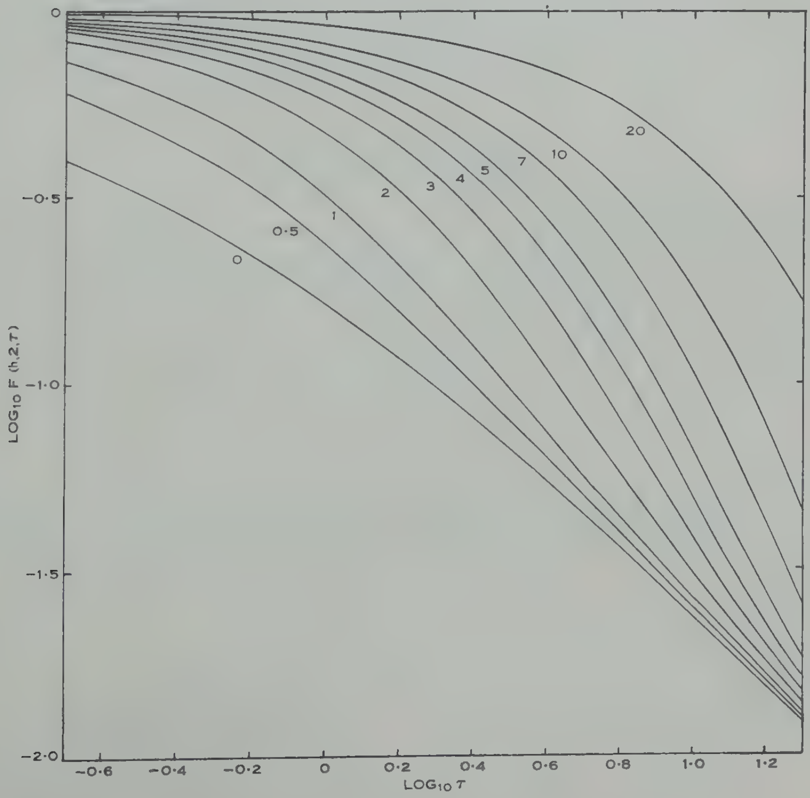


Fig. 2.—The function $F(h, 2, \tau)$. Numbers on the curves are the values of h.

The values of τ in Table 1 and Figures 1 and 2 are those in the region which was shown in Section II to be important for many practical applications. For these values of τ the integral (**6**) must be evaluated numerically. For larger or smaller values, approximations are available.

If τ is small it may be shown as in Carslaw and Jaeger (1947) that

$$F(h,\ \alpha,\ \tau)=1-(\alpha\tau/h)+O(\tau^{3/2}), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{8})$$

except for $h=0$, in which case it becomes

$$F(h,\ \alpha,\ \tau)=1-\frac{2\alpha}{\pi^{\frac{1}{2}}}\tau^{\frac{1}{2}}+\alpha(\alpha-\tfrac{1}{2})\tau+O(\tau^{3/2}). \quad \ldots\ldots\ldots \quad (\mathbf{9})$$

For large values of τ, as in Blackwell (1954),

$$F(h,\ \alpha,\ \tau)=\frac{1}{2\alpha\tau}+\frac{(4h-\alpha)}{4\alpha^2\tau^2}-\frac{(\alpha-2)}{4\alpha^2\tau^2}\left(\ln\frac{4\tau}{C}-1\right)+O(\tau^{-3}\ln\tau), \ldots (\mathbf{10})$$

where

$$C=1\cdot7811\ .\ .\ .=e^{\gamma}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{11})$$

and

$$\gamma=0\cdot5772\ .\ .\ .\ \text{is Euler's constant.}$$

IV. The Temperature in the Solid

In Section III only the temperature in the perfect conductor was discussed. The temperature in the solid involves yet another parameter, r/a, except for its limiting value v_s, as $r\rightarrow a$, which is given by

$$\frac{v_s}{V_0}=F(h,\ \alpha,\ \tau)+\frac{h}{\alpha}\ \frac{\mathrm{d}F(h,\ \alpha,\ \tau)}{\mathrm{d}\tau}. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{12})$$

As remarked in Section I, values of v for $r>a$ may be found from v_s by using Duhamel's theorem and tabulated values for the temperature in the region $r>a$ with its surface maintained at unit temperature.

For the case of perfect thermal contact, $h=0$, only three parameters are involved, and in this case the temperature v for $r>a$ is given by

$$v/V_0=\Phi(\alpha,\ \tau,\ r/a), \quad \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{13})$$

where

$$\Phi(\alpha,\ \tau,\ r/a)=$$
$$\frac{2}{\pi}\int_0^{\infty}\frac{\exp(-\tau u^2)\{J_0(ru/a)[uY_0(u)-\alpha Y_1(u)]-Y_0(ru/a)[uJ_0(u)-\alpha J_1(u)]\}\mathrm{d}u}{[uJ_0(u)-\alpha J_1(u)]^2+[uY_0(u)-\alpha Y_1(u)]^2}.$$
$$\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{14})$$

Some values of $\Phi(\alpha,\ \tau,\ r/a)$ for the most interesting case $\alpha=2$ are shown in Figure 3.

The value of v given by (**13**) may be regarded as the temperature at r due to the liberation of an amount of heat $Q'=SV_0$ per unit length over the cylindrical surface $r=a$. This is the Green's function of the problem, and, by the symmetry of the Green's function or by a direct calculation as in Jaeger (1940), it follows that the temperature V in $r<a$ due to the liberation of an amount of heat Q' per unit length over the cylindrical surface $r=r'$ is

$$V=\frac{Q'}{S}\Phi(\alpha,\ \tau,\ r'/a). \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{15})$$

In view of the complicated nature of the result (**14**) it is of some interest to remark that for the case $\alpha=2$ the values of (**15**) taken from Figure 3 differ very little (except for $r/a=1\cdot5$) from those for the cylindrical surface source (Carslaw and Jaeger 1947, Section 103) in an infinite medium, namely,

$$\frac{Q'}{4\pi Kt}\exp\{-(r'^2+a^2)/4\varkappa t\}I_0\left(\frac{ra}{2\varkappa t}\right). \quad \ldots\ldots\ldots\ldots \quad (\mathbf{16})$$

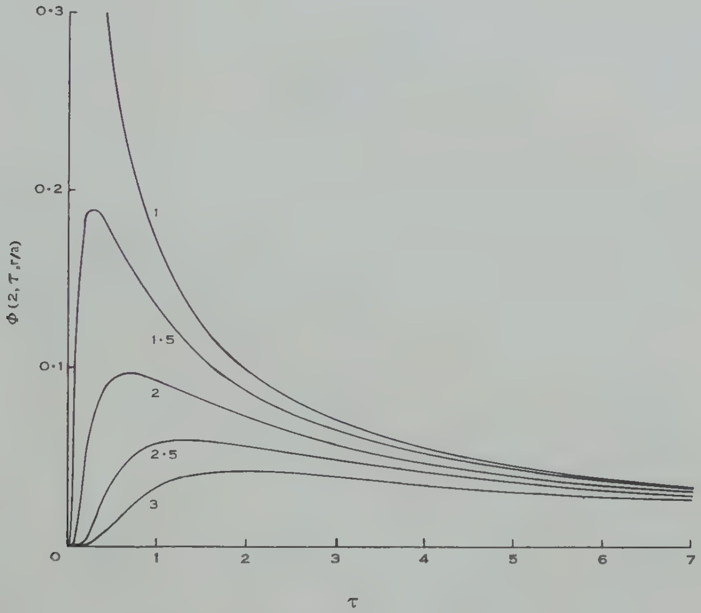


Fig. 3.—Values of $\Phi(2,\ \tau,\ r/a)$. The numbers on the curves are the values of r/a.

V. Constant Supply of Heat in the Perfect Conductor

In this case it is assumed that both the regions $r<a$ and $r>a$ are initially at zero temperature and that heat is supplied at the constant rate Q per unit time per unit length to the perfect conductor. The temperature V of the perfect conductor is given by

$$KV/Q=G(h,\ \alpha,\ \tau), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{17})$$

where

$$G(h,\ \alpha,\ \tau)=\frac{2\alpha^2}{\pi^3}\int_0^\infty\frac{\{1-\exp(-\tau u^2)\}\mathrm{d}u}{u^3\Delta(u)}, \quad \ldots\ldots\ldots \quad (\mathbf{18})$$

and $\Delta(u)$ is given by (**7**).

The quantity

$$SV/Qt=(2\pi/\alpha\tau)G(h,\ \alpha,\ \tau) \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{19})$$

is the ratio of the actual temperature to that which would be attained in the absence of loss by conduction and is useful in the reduction of experimental data.

The temperature for $r>a$ may be found as before. The most interesting quantity is the temperature drop over the contact resistance at $r=a$: this is given by

$$\frac{Qh}{2K\pi}\{1-F(h,\ \alpha,\ \tau)\}, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{20})$$

where $F(h,\ \alpha,\ \tau)$ is defined by (**6**) and was discussed in Section III.

TABLE 2

VALUES OF $G(0,\ \alpha,\ \tau)$

α τ	0·5	1·0	1·5	2·0	4·0	6·0	8·0	∞
0·2	0·013	0·023	0·030	0·035	0·047	0·052	0·056	0·068
0·3	0·019	0·032	0·040	0·046	0·060	0·066	0·069	0·080
0·4	0·025	0·040	0·050	0·056	0·070	0·076	0·079	0·090
0·5	0·030	0·047	0·058	0·065	0·079	0·085	0·088	0·098
0·6	0·035	0·054	0·065	0·073	0·087	0·093	0·096	0·105
0·7	0·039	0·060	0·072	0·080	0·094	0·100	0·102	0·112
0·8	0·044	0·066	0·078	0·086	0·100	0·106	0·109	0·118
0·9	0·048	0·071	0·084	0·092	0·106	0·112	0·115	0·123
1·0	0·052	0·077	0·090	0·098	0·112	0·117	0·119	0·128
2·0	0·087	0·117	0·130	0·138	0·150	0·154	0·156	0·162
3·0	0·114	0·145	0·158	0·165	0·175	0·178	0·180	0·185
4·0	0·136	0·166	0·178	0·184	0·194	0·197	0·198	0·202
5·0	0·155	0·184	0·195	0·200	0·209	0·211	0·213	0·216
6·0	0·170	0·198	0·208	0·214	0·221	0·224	0·225	0·228
7·0	0·184	0·211	0·220	0·225	0·232	0·234	0·235	0·238
8·0	0·196	0·222	0·230	0·235	0·241	0·243	0·244	0·247
9·0	0·207	0·231	0·240	0·244	0·250	0·252	0·252	0·255
10·0	0·217	0·240	0·248	0·252	0·257	0·259	0·260	0·262
15·0	0·255	0·274	0·280	0·283	0·287	0·288	0·289	0·290
20·0	0·282	0·297	0·302	0·305	0·308	0·309	0·310	0·311

For the case of perfect thermal contact at $r=a$, $h=0$,values of $G(0,\ \alpha,\ \tau)$ are given in Table 2. It may be noted that the case $\alpha\rightarrow\infty$ is that of prescribed flux Q at $r=a$ into the region $r>a$ which is treated in Carslaw and Jaeger (1947, Section 127).

For the general case, values of $G(h,\ \alpha,\ \tau)$ are shown in Figures 4 and 5 for $\alpha=1$ and $\alpha=2$, respectively, G being plotted against log τ.

If τ is small,

$$G(h,\ \alpha,\ \tau)=\frac{\alpha}{2\pi}\left\{\tau-\frac{\alpha\tau^2}{2h}+O(\tau^{5/2})\right\}, \quad \ldots\ldots\ldots\ldots \quad (\mathbf{21})$$

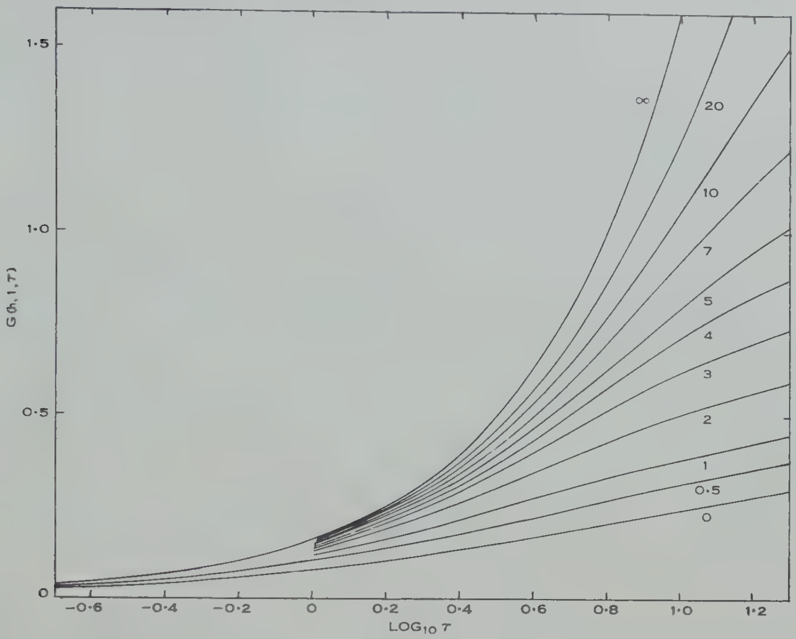


Fig. 4.—The function $G(h, 1, \tau)$. The numbers on the curves are the values of h.

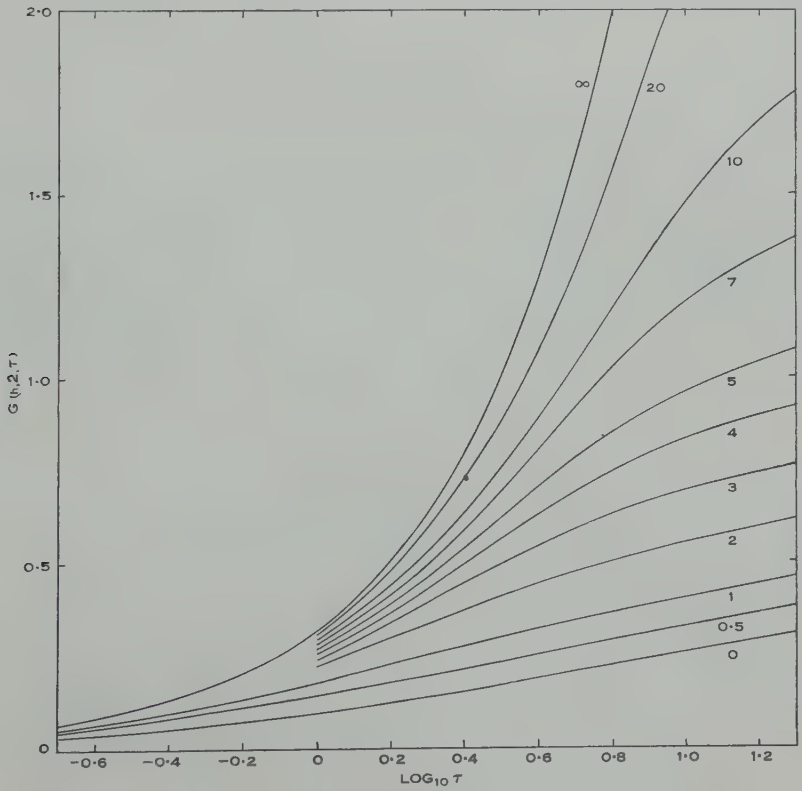


Fig. 5.—The function $G(h, 2, \tau)$. The numbers on the curves are the values of h.

if $h \neq 0$, while if $h=0$,

$$G(0, \alpha, \tau)=\frac{\alpha}{2\pi}\left\{\tau-\frac{4\alpha}{3\pi^{\frac{1}{2}}}\tau^{3/2}+O(\tau^2)\right\}. \quad \ldots\ldots\ldots\ldots (22)$$

For large values of τ,

$$G(h, \alpha, \tau)=\frac{1}{4\pi}\left\{2h+\ln\frac{4\tau}{C}-\frac{(4h-\alpha)}{2\alpha\tau}+\frac{\alpha-2}{2\alpha\tau}\ln\frac{4\tau}{C}+\ldots\right\}, \quad \ldots (23)$$

where C is defined in (**11**). The fact that for large values of τ, G varies linearly with $\ln\tau$ has been much used for the determination of thermal conductivity. It appears from Figures 4 and 5 that for $h>1$ the curve of G against $\ln\tau$ has an inflexion between $\tau=2$ and $\tau=10$ so that, if an experimental curve terminates in this region, this might be regarded as the asymptote (**23**) and a completely wrong value of K found.

It may be remarked that the functions $F(h, \alpha, \tau)$ and $G(h, \alpha, \tau)$ are connected by the relation

$$\frac{d}{d\tau}G(h, \alpha, \tau)=\frac{\alpha}{2\pi}F(h, \alpha, \tau). \quad \ldots\ldots\ldots\ldots\ldots (24)$$

The following limiting relations are also useful.

$$G(\infty, \alpha, \tau)=\alpha\tau/2\pi, \quad \ldots\ldots\ldots\ldots\ldots (25)$$

$$G(h, \infty, \tau)=h/2\pi+G(0, \infty, \tau). \quad \ldots\ldots\ldots\ldots\ldots (26)$$

VI. The Case of Moving Fluid

Another problem which may be solved in the same way is that in which the region $r<a$ contains mass M_1 per unit length of well-stirred fluid of specific heat c_1. Mass m_1 of fluid is removed from the region per unit time per unit length and replaced by fluid at temperature V_0. If both the fluid in the region $r<a$ and the solid in the region $r>a$ are initially at zero temperature, and heat is supplied in the region $r<a$ at the rate Q per unit time per unit length, the temperature V in this region at time t is given by

$$V=\left(V_0+\frac{Q}{m_1c_1}\right)H(\mu, \alpha, \tau), \quad \ldots\ldots\ldots\ldots (27)$$

where

$$H(\mu, \alpha, \tau)=1-\frac{4\alpha\mu}{\pi^2}\int_0^\infty\frac{\exp(-\tau u^2)du}{u\Delta(\mu, \alpha, u)}, \quad \ldots\ldots\ldots (28)$$

$$\Delta(\mu, \alpha, u)=[(\mu-u^2)J_0(u)+\alpha uJ_1(u)]^2+[(\mu-u^2)Y_0(u)+\alpha uY_1(u)]^2, \ldots (29)$$

μ is the dimensionless parameter

$$\mu=m_1a^2/\varkappa M_1, \quad \ldots\ldots\ldots\ldots\ldots (30)$$

and τ and α are defined in (**2**) and (**3**).

Some values of $H(\mu, \alpha, \tau)$ for the case $\alpha=1$ are shown in Figure 6. This theory has been used for the measurement of water movement in or across boreholes ; an account of the method will be given elsewhere.

For large values of the time a first approximation to the value of $H(\mu, \alpha, \tau)$ is

$$1-\frac{4\alpha}{\mu\pi^2}I(0, 1, \tau), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (31)$$

where $I(0, 1, \tau)$ is the function specifying the flux into the region $r>a$ for constant surface temperature which has been tabulated by Jaeger and Clarke (1942).

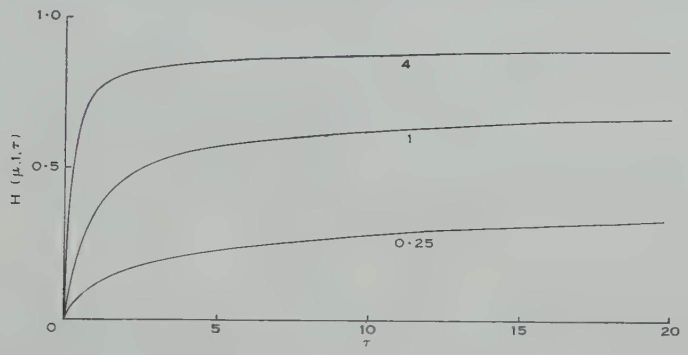


Fig. 6.—The function $H(\mu, 1, \tau)$. The numbers on the curves are the values of μ.

VII. THE HEATING OF BURIED ELECTRIC CABLE

The fundamental situation in the heating of buried cable is stated by Whitehead and Hutchings (1938) as follows : the current-carrying core of the cable consists of a perfect conductor of thermal capacity S_1 per unit length ; this is separated by insulating material from a protecting sheath which consists of perfect conductor of thermal capacity S_2 per unit length ; the external radius of the sheath is a and it is surrounded by earth of thermal constants K, ρ, c, $\varkappa$, there being no contact resistance between the sheath and the surrounding earth. The thermal capacity of the cable insulation is neglected, and the thermal resistance between the core and the sheath will be written in the form $1/(2\pi aH)$ per unit length to conform with the notation used earlier.

The essential problem is that of heat supply in the core at the rate Q per unit time per unit length starting at time $t=0$. The most important practical cases are : (i) when a peak load is suddenly added to steady conditions due to a base load, in which case the time of heating is of the order of an hour, and (ii) the case of a short circuit in which the core is heated almost instantaneously and it is required to find the way in which this heat is dissipated.

Whitehead and Hutchings (1938) discuss the first of these problems, making a number of approximations so that, while their result is valid for very large values of the time, it is quite impossible to estimate its error at moderate times. In connexion with the second problem, Whitehead (1944) gives an approximate solution of the problem of Section III for $h=0$, and Tranter (1947) has pointed out that it is more satisfactory to evaluate the integrals directly than to use such approximations.

The work of the preceding sections applies immediately to the case of the unsheathed cable, $S_2=0$, which, however, is not of any great engineering importance. It also applies to the case in which the thermal capacity of the core is negligible, so that, if heat is supplied at the rate Q per unit time per unit length to the core, it may be regarded as being supplied at the same rate to the sheath through the thermal resistance $1/(2\pi aH)$. The temperature V_2 in the sheath is then given by

$$V_2=\frac{Q}{K}G(0,\ \alpha_2,\ \tau), \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{32})$$

where

$$\alpha_2=2\pi a^2\rho c/S_2, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{33})$$

and the temperature V_1 of the core is given by

$$V_1=V_2+\frac{Q}{2\pi aH}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{34})$$

The assumption leading to (**32**) is that made by Whitehead and Hutchings (1938) in their Appendix III, so that Table 2 provides exact values for their quantity β.

Accurate expressions for the complete system are easily obtained. If the core is initially at temperature V_0 and the rest of the system at zero temperature, no heat being supplied to the core, its subsequent temperature V_1 is given by

$$\frac{V_1}{V_0}=\frac{4\alpha_1\alpha_2^2}{\pi^2}\int_0^\infty\frac{\exp(-u^2\tau)du}{u\Delta_1(u)}, \qquad \ldots\ldots\ldots\ldots \quad (\mathbf{35})$$

where h and τ are defined in (**4**) and (**2**),

$$\alpha_1=\frac{2\pi a^2\rho c}{S_1}, \qquad \alpha_2=\frac{2\pi a^2\rho c}{S_2}, \qquad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{36})$$

and

$$\begin{aligned}\Delta_1(u)=&[u(\alpha_1+\alpha_2-hu^2)\mathrm{J}_0(u)-\alpha_2(\alpha_1-hu^2)\mathrm{J}_1(u)]^2\\&+[u(\alpha_1+\alpha_2-hu^2)\mathrm{Y}_0(u)-\alpha_2(\alpha_1-hu^2)\mathrm{Y}_1(u)]^2. \quad \ldots\ldots \quad (\mathbf{37})\end{aligned}$$

For the case in which the whole system is initially at zero temperature and heat is supplied at the rate Q per unit time per unit length to the core, its temperature V_1 at time t is given by

$$V_1=\frac{2Q\alpha_1^2\alpha_2^2}{K\pi^3}\int_0^\infty\frac{\{1-\exp(-u^2\tau)\}du}{u^3\Delta_1(u)}, \qquad \ldots\ldots\ldots \quad (\mathbf{38})$$

where the symbols are those defined above. The application of these results to engineering practice will be discussed more fully elsewhere.

VIII. ACKNOWLEDGMENT

The author is indebted to Mrs. A. Davidson for most of the detailed numerical work.

IX. REFERENCES

BECK, A., JAEGER, J. C., and NEWSTEAD, G. (1956).—*Aust. J. Phys.* **9** : 286.

BLACKWELL, J. H. (1954).—*J. Appl. Phys.* **25** : 137.

BULLARD, E. (1954).—*Proc. Roy. Soc.* A **222** : 408.

CARSLAW, H. S., and JAEGER, J. C. (1947).—" Conduction of Heat in Solids." (Oxford Univ. Press.)

FISCHER, J. (1939).—*Ann. Phys., Lpz.* **34** : 669.

VAN DER HELD, E. F. M., HARDEBOL, J., and KALSHOVEN, J. (1953).—*Physica, 's Grav.* **19** : 208.

JAEGER, J. C. (1940).—*J. Roy. Soc. N.S.W.* **74** : 342.

JAEGER, J. C. (1956).—*J. Math. Phys.* **34** : 316.

JAEGER, J. C., and CLARKE, MARTHA (1942).—*Proc. Roy. Soc. Edinb.* A **61** : 229.

TRANTER, C. J. (1947).—*Phil. Mag.* **38** : 131.

WHITEHEAD, S. (1944).—*Proc. Phys. Soc. Lond.* **56** : 357.

WHITEHEAD, S., and HUTCHINGS, E. E. (1938).—*J. Instn. Elect. Engrs.* **83** : 517.

THERMAL AND ELECTRICAL CONDUCTIVITIES OF IRON, NICKEL, TITANIUM, AND ZIRCONIUM AT LOW TEMPERATURES

By W. R. G. KEMP,* P. G. KLEMENS,* and G. K. WHITE†

[*Manuscript received November 22, 1955*]

Summary

Measurements are reported of the thermal and electrical conductivities of iron, nickel, titanium, and zirconium down to 2 °K. These indicate that thermal conduction in pure iron and nickel is almost completely electronic. Titanium and zirconium exhibit an appreciable lattice component of thermal conduction. In the case of titanium this lattice component varies as $T^{1\cdot5}$.

The ideal electronic thermal resistivity below about 50 °K was found to be $10\times10^{-5}\,T^{2\cdot2}$ and $11\times10^{-5}\,T^2$ cm deg W^{-1} for iron and nickel respectively. The ideal electrical resistivity was found to vary as T^3 at low temperatures for all four metals.

I. INTRODUCTION

Rosenberg (1955) has described the results of a comprehensive investigation of the low temperature thermal conductivity of a large number of metallic elements. These results confirm the generally accepted view that the thermal conductivity of those metals which are good electrical conductors is principally electronic. The electronic thermal conductivity, $\varkappa_e$, can be expressed as

$$1/\varkappa_e=W_e=W_i+W_0, \qquad \ldots\ldots\ldots\ldots \quad (1)$$

where W_i, the ideal or intrinsic thermal resistivity, arises from scattering by lattice waves and W_0, the residual thermal resistivity, is due to scattering by static imperfections. One would expect from theory, irrespective of the details of the electronic band structure, that

$$W_0=\rho_0/L_0T, \qquad \ldots\ldots\ldots\ldots \quad (2)$$

and, for $T\ll\theta$,

$$W_i=BT^n, \qquad \ldots\ldots\ldots\ldots \quad (3)$$

where $n\simeq2$, ρ_0 is the residual electrical resistivity, and $L_0=2\cdot45\times10^{-8}$ W Ω deg^{-2}, the Sommerfeld value of the Lorenz ratio (see, for example, Klemens 1956).

In the absence of an appreciable lattice component of thermal conduction the Lorenz ratio $L=\rho/WT$ should tend to the value L_0 at those temperatures at which ρ becomes constant, since the electrical and thermal resistivities are then ρ_0 and W_0 respectively. Rosenberg, however, found that in the cases of iron, titanium, and zirconium L deviated markedly from L_0 even at liquid helium temperatures, the observed thermal conductivity being larger than

---

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

† Division of Physics, C.S.I.R.O.; present address: Division of Pure Physics, National Research Council, Ottawa, Canada.

would be expected from the substitution of the observed value of ρ_0 into equation (**2**). He found for titanium and zirconium that the thermal resistance was inversely proportional to temperature and concluded from this that for these metals the observed discrepancy could not be attributed to the presence of an appreciable lattice component of thermal conductivity ($\varkappa_g$) but must be due to a failure of the Lorenz law (**2**).

In view of the very general theoretical validity of (**2**), its failure would be very disturbing. It thus seemed worth while to reinvestigate these materials and to check that any deviations from (**2**) cannot, in fact, be attributed to lattice conduction, especially as recent measurements on silver alloys (Kemp *et al.* 1954, 1956) and on beryllium and copper (White and Woods 1955), as well as theoretical studies (Klemens 1955), indicate that $\varkappa_g$ may be appreciable and show various types of temperature variation, depending upon the nature of the principal lattice imperfections.

The thermal and electrical conductivities of iron and nickel were investigated as an extension of the work on transition metals previously reported (Kemp *et al.* 1955).

II. Experimental Procedure and Specimens

The thermal conductivity measurements were made in a cryostat described previously (White 1953) and the electrical conductivity was measured simultaneously in the same cryostat with the aid of a galvanometer amplifier (MacDonald 1947). The specimens were supplied by Messrs. Johnson, Matthey and Co. Ltd.; Table 1 gives details of their physical state and purity.

Table 1

PURITY AND PHYSICAL STATE OF SPECIMENS

Element	Source Material	Nominal Purity (%)	Analysis	Specimen
Iron	JM5092	99·99	Ni (0·005%), Cu (0·0002%), Ag (0·0001%), Mn, Mg (barely visible lines)	2 mm dia. rod annealed 750 °C for 4 hr *in vacuo*
Nickel . .	JM4497	>99·99	Si, Ca, Al, Ag, Cu (faint lines) Mg, Na, Li (very faint lines)	2 mm dia. rod annealed 750 °C for 4 hr *in vacuo*
Zirconium	JM5000	99·99	Fe (all sensitive lines) Hf, Ni (all sensitive lines faintly) Si, Ti (some sensitive lines) Cr, Al, Cu, Mg (faintly visible)	3 mm dia. rod annealed 950 °C for 5 hr *in vacuo*
Titanium . .	JM4233	98	Mg (0·024%), Si (0·13%), Fe (0·05%), Ni (0·081%), C (0·14%), O_2 (1·63%)	3 mm dia. rod annealed 950 °C for 5 hr *in vacuo*

Connections to the iron and nickel specimens were made with a zinc-cadmium eutectic solder which is not superconducting down to 2 °K, the temperature range of the measurements.

Initially the specimens of zirconium and titanium were mounted in the cryostat by means of copper fittings, into which the specimens were a push fit, and were sealed with baked "Araldite" cement. These were designated Zr1a and Ti1a. However, as the electrical resistance measurements were not reproducible, and on examination the contacts provided by the copper rings were found to be unsatisfactory, an alternative mounting was adopted for subsequent measurements. Holes were drilled and tapped in the zirconium and titanium specimen rods and current and potential connections (both electrical and thermal) made by means of 10 B.A. screws, which were tightly screwed into the specimens (designated Zr1c and Ti1b).

After the initial measurements on Zr1a were completed, it was thought that the current lead used for the measurements of electrical resistance may have affected the apparent thermal conductivity, as this lead was a 44 S.W.G. copper wire (later replaced by 34 S.W.G. constantan). The thermal conductivity was therefore measured again with the current lead removed (Zr1b); the values obtained nowhere differed by more than 3 per cent. from those for Zr1a.

III. Results

(*a*) *Iron*

While the values of the thermal conductivity (Fig. 1) appear to agree quite well with those found by Rosenberg (1955) on his first run, values of WT and ρ are both constant to within 2 per cent. below 15 °K. Rosenberg found on his second run, in which he measured both electrical and thermal conductivities, a small anomaly in WT below 1·0 °K and a sharp drop in ρ of about 10 per cent.

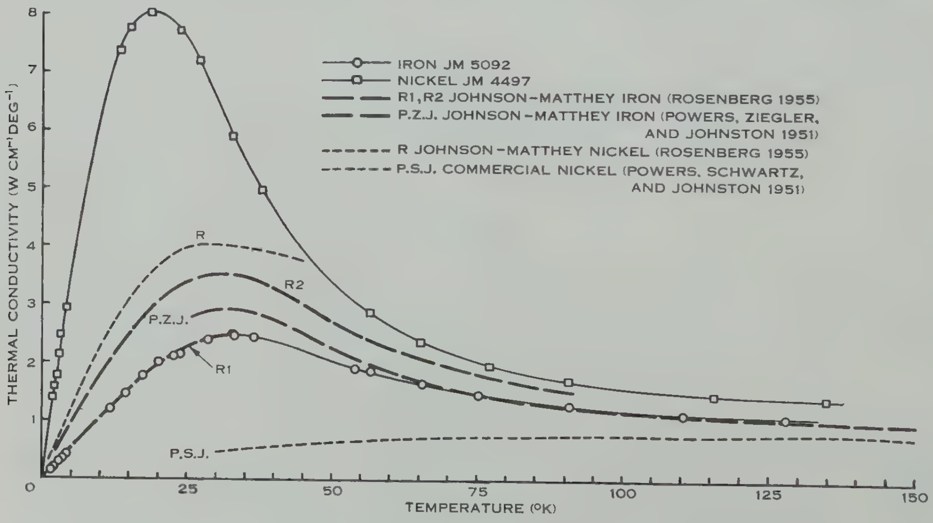


Fig. 1.—Thermal conductivity of iron and nickel at low temperatures.

around 8 °K; his Lorenz ratio has a sharp peak at about 8 °K, and below this it falls from $2{\cdot}6$ to $2{\cdot}3\times10^{-8}$ W Ω deg^{-2}, whereas our specimen (Fig. 2) has

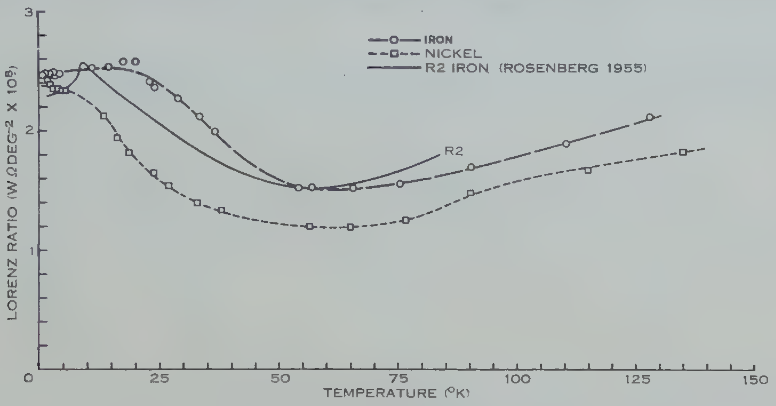


Fig. 2.—Variation of Lorentz ratio $\rho\varkappa/T$ with temperature.

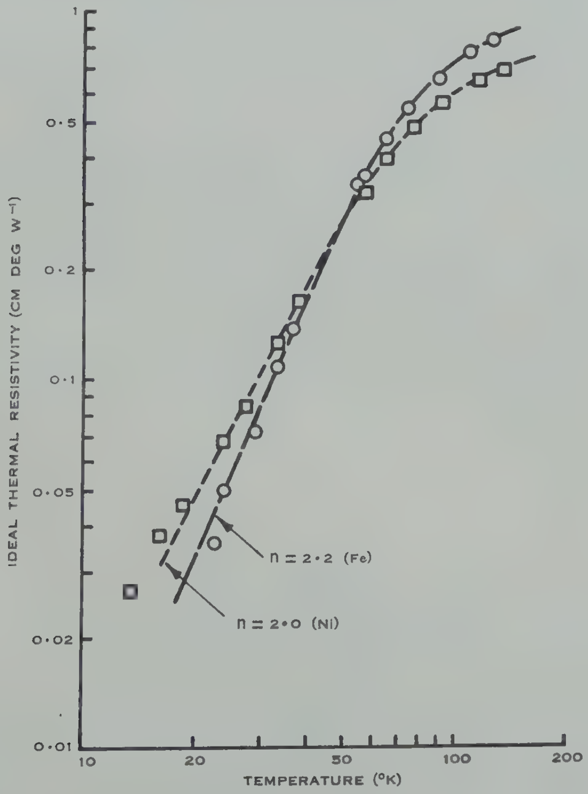


Fig. 3.—Ideal thermal resistivity of iron and nickel at low temperatures.

a value of $L=\rho/WT$ which is fairly constant (about $2{\cdot}5\times10^{-8}$) from 20 °K down.

It seems possible that the anomaly in the electrical resistance of Rosenberg's iron specimen was due to the solder (Wood's metal or soft solder) becoming superconducting below about 8 °K.

The values of the ideal thermal resistivity W_i, obtained from (**1**) by assuming the experimental figure $W_0T=10\cdot0$ cm deg^{-2} W^{-1}, are plotted in Figure 3 and indicate that equation (**3**) is satisfied below 40 °K ($\theta/10$) with $n\simeq2\cdot2$. If we assume $n=2$ and calculate B at 30 °K, the value 10×10^{-5} cm deg^{-1} W^{-1} is obtained, in close agreement with Rosenberg's values of 10·2 (R1) and 9·5 (R2).

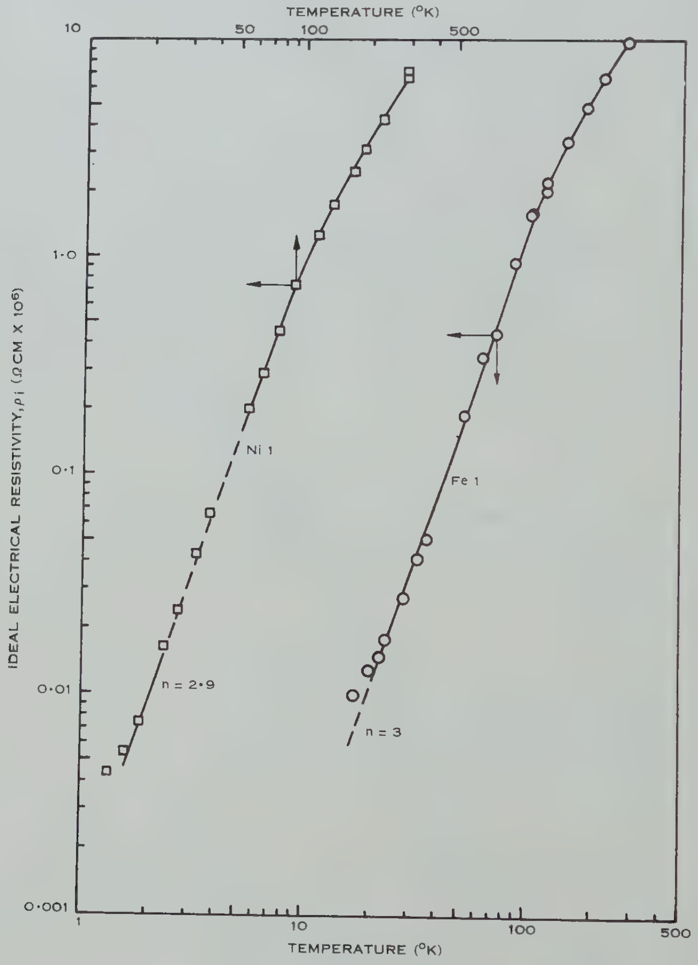


Fig. 4.—Ideal electrical resistivity of iron and nickel.

The electrical resistivities at 20 °C and at helium temperatures are respectively $10\cdot0\times10^{-6}$ and $0\cdot248\times10^{-6}$ Ω cm, so that $\rho_0/\rho_{293}=2\cdot48\times10^{-2}$. The ideal electrical resistivity $\rho_i=\rho-\rho_0$ is shown in Figure 4. Below 90 °K $\rho_i\simeq1\cdot25\times10^{-12}T^3$ Ω cm.

(*b*) *Nickel*

The thermal conductivity, Lorenz ratio, and ideal thermal and electrical resistivities of Ni1 are shown in Figures 1–4. Both ρ and WT are sensibly constant in the liquid helium region, and the Lorenz ratio L has a value of $2 \cdot 38 \times 10^{-8}$, which is not inconsistent with the theoretical value, since the possible errors in the electrical and thermal measurements are both in the vicinity of 1 per cent. The values of ρ_{293} and ρ_0 are $7 \cdot 22 \times 10^{-6}$ and $0 \cdot 0347 \times 10^{-6}\ \Omega$ cm respectively, and the low value of $\rho_0/\rho_{293} \simeq 4 \cdot 8 \times 10^{-3}$ indicates the high purity of the specimen. Below about 40 °K, $W_i = 11 \times 10^{-5} T^2$ deg cm W^{-1} (see Fig. 3), as compared with Rosenberg's value of $10 \cdot 4 \times 10^{-5}$ for B: Figure 4 indicates that $\rho_i = 1 \cdot 19 \times 10^{-12} T^3\ \Omega$ cm below about 90 °K.

(*c*) *Titanium*

The thermal conductivities and electrical resistivities of titanium and zirconium are shown in Figures 5 and 6. The relatively impure titanium specimen has a smaller thermal conductivity than Rosenberg's specimens. The values of ρ_{293} and ρ_0 are 70×10^{-6} and $23 \cdot 6 \times 10^{-6}\ \Omega$ cm respectively. Below 50 °K ($\sim\theta/5$), $\rho_i \simeq 17 \times 10^{-12} T^3\ \Omega$ cm; however, the ideal resistance is not determined accurately, since for such an impure specimen $\rho_0 \gg \rho_i$ at low temperatures.

This high residual resistance also makes it impossible to deduce values of W_i at low temperatures. The value of $4 \cdot 5 \times 10^{-3} T^2$ for W_i, obtained by Rosenberg (1955) for a pure single crystal of titanium ($\rho_0 = 2 \cdot 4 \times 10^{-6}$), confirms that W_i will be negligible in comparison with W_0 below 40 °K for the present specimen. If the lattice component of thermal conductivity were negligible at these temperatures, $\varkappa/T$ should be constant and, from (**2**), equal to L_0/ρ_0. The measurements indicate that $\varkappa/T$ is not constant, and is larger than L_0/ρ_0. This suggests that there is an appreciable heat transport by the lattice. If this is so, then the lattice thermal conductivity should be given by

$$\varkappa_g = \varkappa - \varkappa_e = \varkappa - L_0 T/\rho_0, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{4})$$

and, when thus obtained, can be expressed as $\varkappa_g = 1 \cdot 8 \times 10^{-4} T^{1 \cdot 5}$ W cm^{-1} deg^{-1} between 2 and 30 °K.

(*d*) *Zirconium*

Figure 5 illustrates the thermal conductivity of the zirconium samples 1a, 1b, and 1c. The conductivity of 1c is about 10 per cent. lower, apparently due to strains induced by drilling and tapping to insert the connectors for mounting. In the case of titanium this produced no noticeable change because its residual resistivity was already very high.

Equations (**1**), (**2**), and (**3**) seem reasonably well obeyed, and it is found that in (**3**) $n \simeq 2$ and $B = 1 \cdot 3 \times 10^{-3}$ cm deg^{-1} W^{-1}. The value for B is in good agreement with the values obtained by Rosenberg. The electrical resistivities ρ_{293} and ρ_0 are 48×10^{-6} and $1 \cdot 98 \times 10^{-6}\ \Omega$ cm respectively and ρ_i varies approximately as T^3 below 40 °K. In contrast to Rosenberg's result, the Lorenz ratio

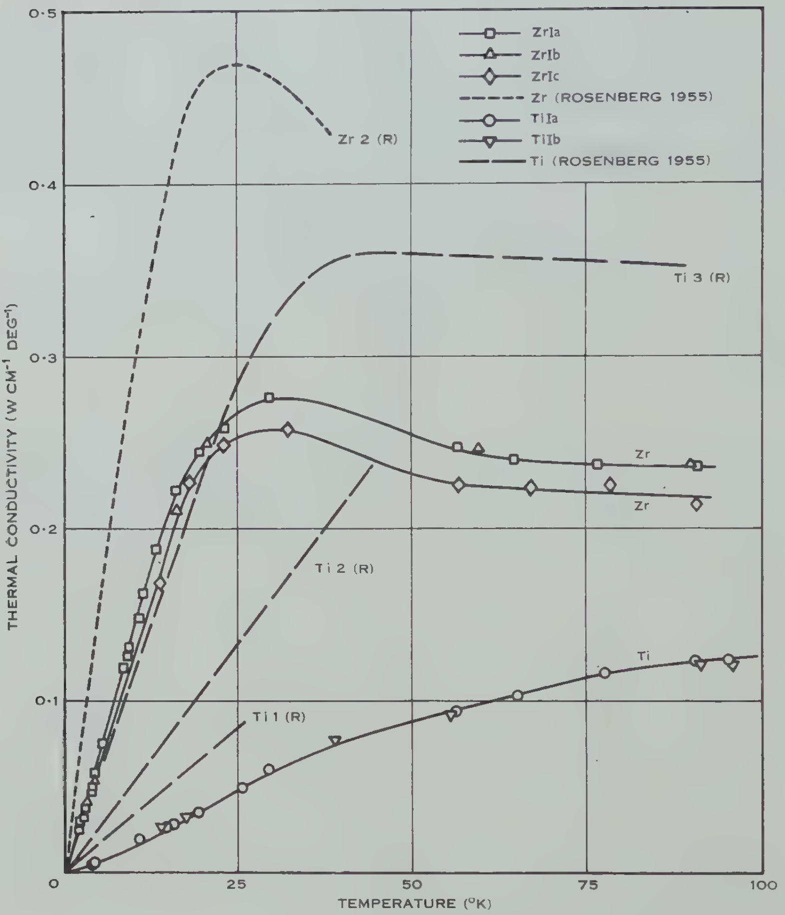


Fig. 5.—Thermal conductivity of titanium and zirconium at low temperatures.

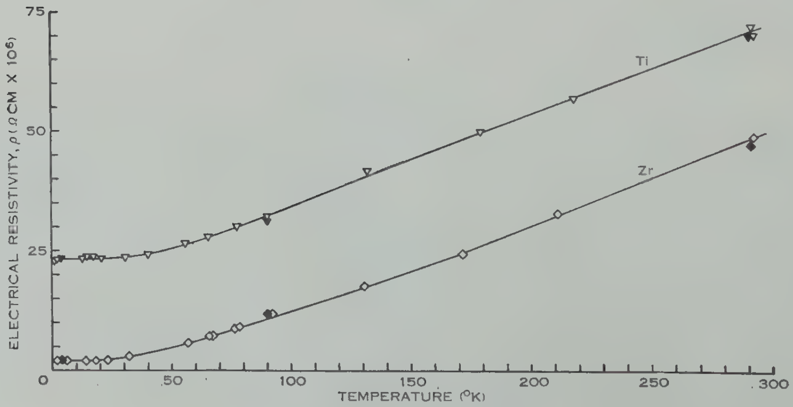


Fig. 6.—Electrical resistivity of titanium and zirconium.

has a sensibly constant value of $2 \cdot 4 \times 10^{-8}$ W Ω deg^{-2} below 25 °K (see Fig. 7). It falls to a slightly lower value at about 40 °K and then increases again with increasing temperature to well above L_0. Such behaviour is consistent with the presence of an appreciable lattice component of thermal conductivity; however, it is not possible to deduce its value from these results.

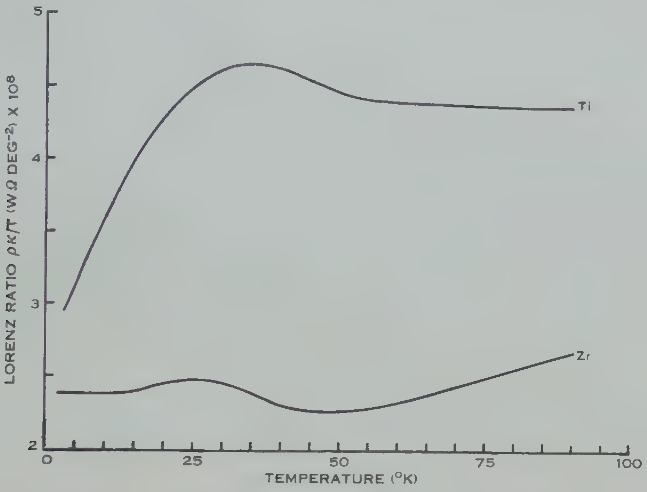


Fig. 7.—Lorenz ratio $\rho\varkappa/T$ for titanium and zirconium.

IV. Discussion

It appears that the electronic conduction properties of iron and nickel can be interpreted along the same lines as those of palladium (Kemp *et al.* 1955). The conductivities are mainly due to electrons in the s-band, and are limited by $(s\text{–}s)$ and $(s\text{–}d)$ transitions. The resistances due to $(s\text{–}s)$ transitions are described by the Bloch single-band theory, so that at low temperatures their contribution to the ideal electrical and thermal resistivities, $\rho(s,s)$ and $W(s,s)$, vary as T^5 and T^2 respectively. The $(s\text{–}d)$ transitions can be described by a relaxation time, so that $\rho(s,d)/TW(s,d)=L_0$; it can also be shown that $\rho(s,d)\propto T^3$, since the resistance arises from single-step phonon processes, instead of multiple-step diffusion over the Fermi surface; similarly $W(s,d)\propto T^2$. Thus $L_i=\rho_i/W_iT$ tends to a constant value at lowest temperatures, and in the limit $L_i/L_0=W(s,d)/[W(s,d)+W(s,s)]$. The present results indicate that for iron and nickel $W(s,d)/W(s,s)\simeq 0 \cdot 8$, while for palladium this ratio had been found to be about $0 \cdot 5$.

The electronic conduction properties of titanium and zirconium may also be interpreted similarly, even though the band structure is probably different and more complicated. One can distinguish between transitions within a band (type I) and transitions from one band to another (type II). It again appears that one band only contributes appreciably to the conduction properties. At sufficiently low temperatures $\rho(\mathrm{I})\propto T^5$, $\rho(\mathrm{II})\propto T^3$, $W(\mathrm{I})$ and $W(\mathrm{II})$ both vary as T^2 and $\rho(\mathrm{II})/TW(\mathrm{II})=L_0$. Thus $L_i/L_0=W(\mathrm{II})/[W(\mathrm{I})+W(\mathrm{II})]<1$. In the case of titanium, using Rosenberg's value of W_i and the present results for ρ_i, L_i/L_0

tends to a limit of $\sim 0 \cdot 15$, while for zirconium in the limit of lowest temperatures $L_i/L_0 \sim 0 \cdot 75$.

The present results on titanium and zirconium can, we believe, be explained in terms of an appreciable lattice component of thermal conductivity. They also reveal that the features which led Rosenberg (1955) to reject lattice thermal conduction as responsible for his large values of ρ/WT are not a general property of these substances. He found the values of ρ/WT in the residual resistance region to be (*a*) temperature independent, (*b*) greater than L_0, and (*c*) apparently independent of ρ_0; none of these findings is confirmed by the present results.

Rosenberg's results also can be interpreted in terms of an appreciable lattice component $\varkappa_g$. In his case $\varkappa_g$ would be larger than in the present specimens; this is not unlikely in view of the known sensitivity of $\varkappa_g$ to some lattice imperfections, especially dislocation arrays. In the present titanium specimen the high oxygen content could easily be responsible for a reduction in $\varkappa_g$. In Rosenberg's specimens $\varkappa_g$ would need to be proportional to T. Such a temperature dependence is by no means impossible; it would, for example, result from the scattering of lattice waves by thin sheets of disordered material embedded in a crystal of fixed orientation (Klemens 1956). The apparent independence of ρ/WT of ρ_0 would indicate that the same imperfections which limit $\varkappa_g$ are also responsible for the major part of the residual resistance.

V. References

Kemp, W. R. G., Klemens, P. G., Sreedhar, A. K., and White, G. K. (1954).—*Proc. Phys. Soc. Lond.* A **67**: 728.

Kemp, W. R. G., Klemens, P. G., Sreedhar, A. K., and White, G. K. (1955).—*Phil. Mag.* **46**: 811.

Kemp, W. R. G., Klemens, P. G., Sreedhar, A. K., and White, G. K. (1956).—*Proc. Roy. Soc.* A **233**: 480.

Klemens, P. G. (1955).—*Proc. Phys. Soc. Lond.* A **68**: 1113.

Klemens, P. G. (1956).—Thermal Conductivity of Solids at Low Temperatures. *In* "Encyclopedia of Physics." Vol. 14, Pt. 4. (J. Springer: Heidelberg.) (in press).

MacDonald, D. K. C. (1947).—*J. Sci. Instrum.* **24**: 232.

Powers, R. W., Schwartz, D., and Johnston, H. L. (1951).—Ohio State University Cryogenics Laboratory TR264–5.

Powers, R. W., Ziegler, J. B., and Johnston, H. L. (1951).—Ohio State University Cryogenics Laboratory TR264–6.

Rosenberg, R. M. (1955).—*Phil. Trans. Roy. Soc.* A **247**: 441.

White, G. K. (1953).—*Aust. J. Phys.* **6**: 397.

White, G. K., and Woods, S. B. (1955).—*Canad. J. Phys.* **33**: 58.

THE HEAT CAPACITY OF COPPER BELOW 4·2 °K

By J. A. Rayne*

[*Manuscript received January* 25, 1956]

Summary

The heat capacity of copper has been measured below 4·2 °K with a calorimeter using a mechanical heat switch. The effects of changes in the chemical and physical state of the copper have been investigated with a view to understanding discrepancies between the results of previous workers.

It is found that the γ and θ values for copper of spectroscopic purity are in good agreement with those of Corak *et al.* (1955). Experiments on other specimens, however, reveal a marked dependence of the electronic heat capacity on purity. Possible reasons for this behaviour are discussed, and an evaluation of previous work is given in the light of these results.

I. Introduction

At sufficiently low temperatures, the heat capacity of a normal metal can be expressed in the form

$$C=\gamma T+A\left(\frac{T}{\theta}\right)^3, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where $A=(12\pi^4/5)R$, R being the molar gas constant. In this expression the first term represents the electronic contribution, γ being given by the relation

$$\gamma=\frac{\pi^2}{3}k^2\left(\frac{\mathrm{d}N}{\mathrm{d}E}\right)_{E=E(0)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

where $(\mathrm{d}N/\mathrm{d}E)_{E=E(0)}$ is the density of states at the top of the Fermi surface. The second term represents the lattice contribution, it being supposed that $T\ll\theta$ so that only long-wavelength phonons make an appreciable contribution to the lattice entropy. For many metals this condition is fulfilled in the liquid helium region.

A number of workers (Keesom and Kok 1936 ; Estermann, Friedberg, and Goldman 1952 ; Corak *et al.* 1955) have measured the heat capacity of copper in the liquid helium region. When fitted to equation (1), however, these determinations give a range of values for γ and θ considerably in excess of the estimated errors. Possible reasons for the discrepancies are :

(1) Differences between the vapour pressure-temperature tables used in calibrating the calorimeter thermometers,

(2) experimental errors and in particular effects due to exchange gas adsorbed on the surface of the specimen,

(3) effects due to differences in the chemical or physical states of the specimens.

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

To date it has been impossible to evaluate the experiments on the basis of these possibilities, since the magnitude of the last-mentioned effect was not known. For this reason the present work was undertaken.

II. Experimental Details

(*a*) *Cryostat*

The apparatus is shown schematically in Figure 1. Liquid helium at atmospheric pressure is contained in vessel *A* and is thermally shielded by liquid nitrogen in the outer Dewar *B* and in the radiation shield *C*. By means of the needle valve *E*, liquid can be admitted to the chamber *F* which forms part of the inner can *G*. The space *H* between this can and the outer container *D* is evacuated, and, by controlled pumping on the liquid in *F*, *G* can be kept at any temperature in the range 1·5–4·2 °K. The control is effected by means of a glass Cartesian manostat (Gilmont 1951) in the pumping line, temperatures being determined from the vapour pressure of the liquid using either a mercury or butyl phthallate manometer.

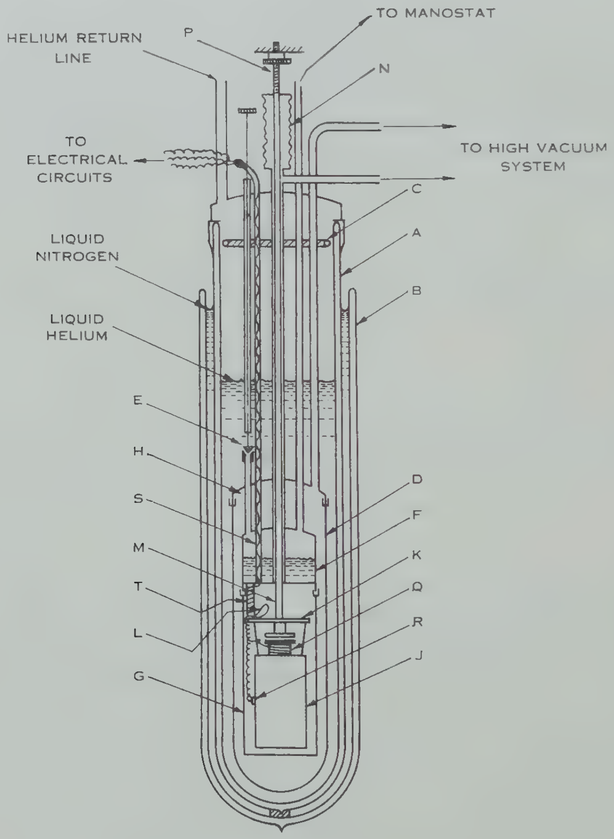


Fig. 1.—Schematic diagram of cryostat.

The calorimeter is similar to that described by Ramanathan and Srinivasan (1955), except that the working space within the inner can is continuously evacuated. Provision is made, however, for admitting exchange gas to permit calibration of the resistance thermometer at the conclusion of an experiment.

Within the working space, the specimen *J* is suspended by nylon threads from the disk *K*, which is thermally anchored to the bath *F* by the flexible copper strap *L*. Movement of the disk is effected by means of a stainless steel tube *M* connected to a bellows *N*, which can be actuated by the screw mechanism *P*. Thermal contact between the specimen *J* and the can *G* is obtained by lowering the disk and pressing the specimen against the flat bottom of the can. When the disk is raised so that the specimen is hanging freely the only heat leak to the specimen arises from conduction down the thread and the electrical connections, stray radiation, and vibration. This heat leak is kept below 100 erg/min.

The heater for the specimen is a manganin resistance of approximately 5000 Ω, which is wound on a copper former *Q* screwed into the top of the specimen. A 10 Ω Allen-Bradley carbon resistance, which is mounted in a copper sleeve *R* screwed into the side of the specimen, is used as a thermometer. Electrical connections to both the heater and thermometer are made with fine manganin wires, which are sealed into and pass down the fine German silver tube *S*. These wires are wound several times around the copper rod *T* to effect thermal contact with the helium bath and hence to reduce the heat leak to the specimen.

(*b*) *Electrical Circuits*

A 1·5 V dry cell supplies energy to the heating circuit, the current being stabilized by a dummy load during the quiescent periods. Switching from the dummy load to the heater is effected by a Leach relay which is actuated by a timer capable of giving a predetermined time interval correct to 1 msec. The heater power is determined by measuring the potential differences across the heater and a 2000 Ω standard resistance on a Diesselhorst potentiometer.

This potentiometer is used to determine the thermometer resistance in the same way. The thermometer circuit is fed from a 6 V accumulator through an 0·6 MΩ resistance in order to minimize variations in the current arising from changes in the thermometer resistance. The current is maintained at approximately 10 μA ; the resulting power input never exceeded 25 erg/min.

(*c*) *Thermometer Calibration*

In the liquid helium region the thermometer characteristics can be approximated by the formula

$$\sqrt{(\log_{10} R/T)}=0\cdot770 \log_{10} R-0\cdot746. \qquad \ldots\ldots\ldots \quad (3)$$

This expression is of the form proposed by Clement, Logan, and Gaffney (1955).

Defining T_R as the value of T corresponding to a given value of R in the above expression, values of $\Delta=T-T_R$ were determined experimentally for a number of temperatures throughout the range 1·5–4·2 °K using the 1949 vapour pressure-temperature tables (Van Dijk and Schoenberg 1949) as corrected by Erickson and Roberts (1954). From these results a correction curve relating Δ to T_R was constructed which, together with equation (3), was used to relate measured values of R to the temperature T.

(*d*) *Specimens*

Three specimens were used in these experiments, namely :

Specimen I—spectroscopically pure copper ;
Specimen II—commercially pure copper ;
Specimen III—copper-1 per cent. cadmium.

Specimen I, obtained from Johnson, Matthey & Co. Ltd., was prepared from material of 99·999 per cent. purity. Measurements on this were made in the " as received " condition. Specimen II was machined from commercial high conductivity, oxygen-free copper of quoted 99·9 per cent. purity. Spectroscopic analysis revealed traces of iron, nickel, and lead, but no estimate of the more volatile impurities is available. Measurements were made on this specimen in both the " as received " and annealed (400 °C, 6 hr) conditions. Specimen III was machined from an ingot of commercial alloy with a nominal cadmium content of 1 per cent. Chemical analysis gave the actual composition as 0·7 per cent. cadmium, 0·5 per cent. zinc. This specimen was measured in the " as received " condition.

(*e*) *Measurements*

Heat capacity measurements were made in the usual manner by following the natural warm-up rate of the specimen and by applying known heat inputs at appropriate intervals. By extrapolating the warm-up curve to the mid points of the heating intervals, the temperature rise ΔT corresponding to a given heat input ΔQ was found. From this the heat capacity at temperature $T+\frac{1}{2}\Delta T$ was calculated from the formula

$$C=\Delta Q/\Delta T. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Corrections for the heat capacity of the heater and thermometer were determined by measuring the heat capacity of two specimens of commercially pure copper having different masses.

III. RESULTS

The experimental results are presented in Figures 2 (*a*) and 2 (*b*) as plots of C/T versus T^2. From equation (1) it can be seen that such a plot should give a straight line whose ordinate at $T=0$ is the value of γ and whose slope is related to the Debye temperature of the material. In each case, it has been possible to fit a straight line to the results with a maximum deviation of any point from the line of less than 2 per cent.

The resulting values of γ and θ, as determined by the method of least squares, are given in Table 1 together with the data obtained by previous workers.

The error estimates for the present work include both random errors, corresponding to the scatter of the experimental data around the lines of best fit, and systematic errors. The latter arise from a number of sources, namely :

(i) Errors in the measurement of heating power arising from the potentiometer and its associated circuit, from the finite resistance of the heater connections, and from drift in the dry cell voltage. These do not exceed 0·1 per cent.

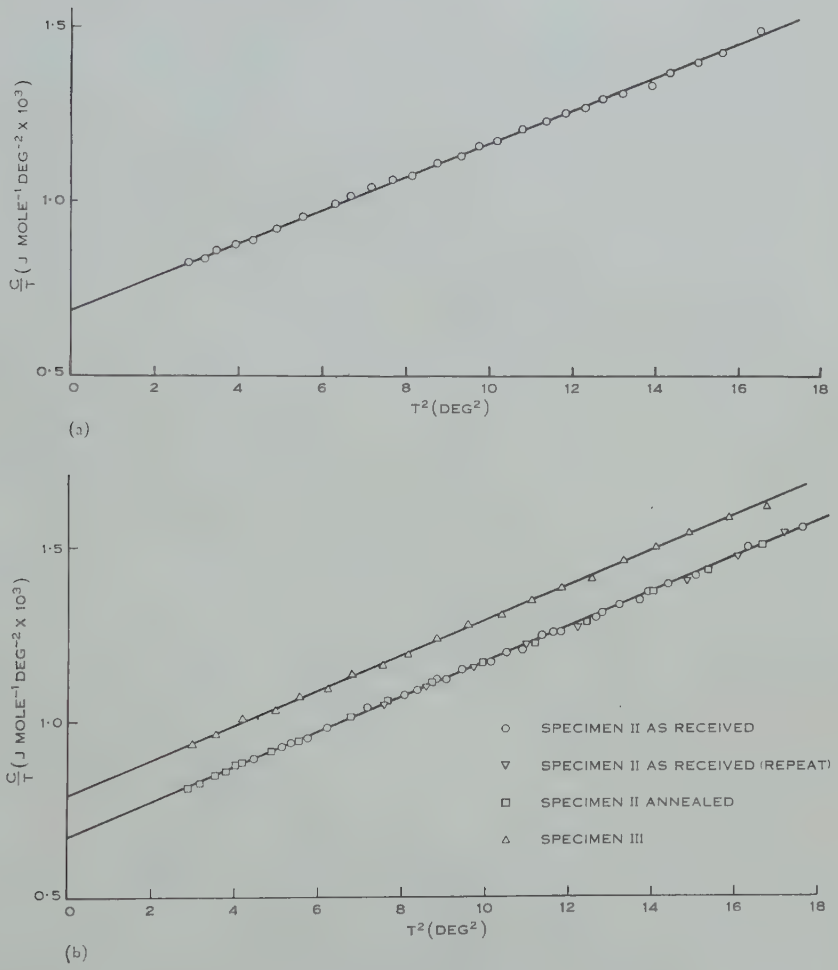


Fig. 2.—Atomic heat of copper. (*a*) Specimen I ; (*b*) specimens II and III.

(ii) Errors in the measurement of heating intervals caused by delays in the timer and actuating relay. These amount to about 4 msec and give an error less than 0·05 per cent.

(iii) Errors in the calibration curve of the thermometer arising from the potentiometer, from the vapour pressure measurements, and from the vapour pressure-temperature relations themselves. The systematic errors resulting from these causes affect the values of mean temperatures, the average uncertainty being 2 millidegrees which corresponds to an error of less than 0·1 per cent. Local errors in slope affect the values of ΔT and hence the heat capacity, but these tend to change sign a number of times and so are effectively included in the estimate of random error.

(iv) Errors in calculating the heat capacity using finite temperature increments ΔT. It can be shown that this error is less than $\frac{1}{4}(\Delta T/T)^2$, which is kept below 0·1 per cent.

(v) Errors in the corrections for the heat capacity of the thermometer and heater. The uncertainties are approximately 5 per cent., but since the correction is only 5 per cent. of the total heat capacity, the resulting error in the latter is only 0·25 per cent.

TABLE 1

VALUES OF γ AND θ

Worker	Specimen and Purity (%)	γ (J mole^{-1} deg$^{-2}\times 10^3$)	θ (°K)
Present author	I, 99·999	0·686±0·005	345·1±0·9
,, ,, ...	II, 99·9	0·670±0·005	338·9±0·9
.. ,, ...	II, 99·9 annealed	0·670±0·005	338·9±0·9
,, ,, ...	III, 99	0·799±0·005	344·7±0·9
Keesom and Kok	99·9	0·74 ±0·02*	335 ±4*
Estermann, Friedberg, and Goldman	99·6	0·75 ±0·01*	315 ±2*
Corak *et al.*	99·999	0·688±0·002	343·8±0·5

* Error estimated by present author from plots of C/T versus T^2.

In all it is estimated that the systematic errors in the value of γ are less than 0·5 per cent., the overall accuracy being approximately 0·8 per cent. The error estimates of Corak *et al.* (1955) given in Table 1 refer only to random errors. When corrected for the systematic error, their overall accuracy is substantially the same as that stated for the present work. No attempt has been made to include any estimate of systematic errors in the other results, since insufficient data are given to assess their magnitude.

IV. DISCUSSION

Reference to Table 1 shows that the present results for material of 99·999 per cent. purity agree with those of Corak *et al.* within the combined limits of systematic and random error, thereby confirming the accuracy of their work and the present work. Since the specimens came from different sources, it is unlikely that they would be identical as regards grain size, state of anneal, etc. From this, together with the fact that specimen II gave identical results in both the strained and annealed states, one can conclude that the values of γ and θ for copper are unaffected by its physical state at least to the accuracy of the present experiments. It would also appear that the value of the Debye temperature is relatively insensitive to impurities.

The results for specimens II and III reveal a marked dependence of electronic heat capacity on impurity content. Such a dependence cannot be understood on the basis of currently accepted theory. According to this, the band structure of copper consists of a narrow, fully occupied *d*-band overlapped by a wide, half filled *s*-band, the density of states being roughly that of a free electron gas with one electron per atom. Impurity atoms, if present, are presumed to contribute their outermost electrons to the collective band system without materially altering its shape, so that a divalent impurity would shift the Fermi level in the direction of increasing energy. Owing to the small slope of the density-of-states curve, however, the resulting change in $(\mathrm{d}N/\mathrm{d}E)_{E=E(0)}$ and hence γ would be quite small for a small impurity concentration. In fact, for the free electron approximation, the relative change in γ is proportional to the relative impurity concentration, the constant of proportionality being of the order unity.

These conclusions are obviously at variance with the present results. One possible explanation for the discrepancy is that the changes in γ are due to changes in the band structure arising from distortions of the lattice caused by impurity atoms. Clearly the magnitude of the effect would depend on the nature of the solute atoms. Elements like cadmium, having a low limit of solid solubility in copper, are known to produce relatively large distortions in the parent lattice and hence one would expect a large effect in this case even though the actual number of impurity atoms was small.

Another possible explanation is to suppose that the actual band structure for copper, in the vicinity of the Fermi level, is a rapidly varying function of energy. Under these conditions relatively small changes in the concentration of valence electrons would produce considerable changes in the density of states and hence γ. Such a situation is not improbable, since in the calculation of the band shape of copper (Krutter 1935) only linear combinations of $3d$ and $4s$ ion core functions were used to compute the metallic wave functions. Actually $4p$ states should also be considered and it is possible that the corresponding $4p$ band overlaps the $4s$ band near the Fermi level. Such a situation would produce an energy band of the required shape to explain the present results. It is of interest to note that Klemens (1954), in explaining the low-temperature conduction properties of the noble metals, has postulated that their Fermi surfaces touch the zone boundaries. Such a situation would also produce a strong energy dependence of the density of states. Further experiments are planned to distinguish between these possibilities.

From the present work, it is possible to draw some conclusions regarding the experiments of Keesom and Kok (1936) and Estermann, Friedberg, and Goldman (1952). In both cases the scatter of points in the plots of C/T versus T^2 is considerable and no substantial improvement results from applying corrections for the errors in the vapour pressure tables used in the corresponding thermometer calibrations. The values of γ and θ still are as good a representation of the data as can be obtained, but the disagreement with the values for pure copper is well outside the combined limits of error, and hence must be due to a combination of experimental error and effects arising from differences in the chemical states of the specimens. In the case of Keesom and Kok, the possible change in γ resulting from the use of a lower purity specimen has been shown here to be at most 3 per cent. Even allowing for this, the difference from the present value for pure copper is outside the limit of error. It would thus seem that at least part of the discrepancy must be attributed to experimental error, the most likely cause arising from the use of exchange gas. The situation regarding the work of Estermann, Friedberg, and Goldman is somewhat more complicated. From the results for the copper-1 per cent. cadmium specimen it would seem that impurity effects could indeed account for the observed discrepancy. It must be remembered, however, that the principal impurity in their sample was lead and that this is virtually insoluble in copper. Apart from the purely bulk effect of lead, the electronic properties of the aggregate would thus be influenced by an extremely small number of atoms and these could scarcely produce the effect observed. Thus the discrepancy must again be due to experimental error although its cause in this case is open to some doubt. There is reason to believe that in part it is due to the neglect of corrections for heat capacity of the heater and thermometer assembly. Such an omission would give too high a value of γ and too low a value of θ. The fact that their value of θ is about 30 °K too low lends considerable weight to this viewpoint, the results of the present work having shown that θ is relatively insensitive to impurity concentration. It appears, however, that the disagreement between the corrected value of γ and the figure for pure copper is still outside the possible limits of error. Thus effects due to exchange gas must also have been present, though their magnitude is difficult to assess.

V. Conclusions

The present work thus substantiates the work of Corak *et al.* (1955) on the low-temperature heat capacity of copper. It shows also the marked dependence of this property on impurity content and allows a critical evaluation of the work of previous observers.

VI. Acknowledgments

Thanks are due to Mr. W. R. G. Kemp for his considerable help in carrying out this work and also to Mr. J. W. Smyth for providing the liquid helium used in these experiments. The author would like also to acknowledge several helpful talks with Dr. Klemens on the interpretation of the results obtained.

VII. REFERENCES

Clement, J. R., Logan, J. K., and Gaffney, J. (1955).—Rep. U.S. Nav. Res. Lab. No. 4542.

Corak, W. S., Garfunkel, M. P., Satterthwaite, C. B., and Wexler, A. (1955).—*Phys. Rev.* **98**: 1699.

Erickson, R. A., and Roberts, L. D. (1954).—*Phys. Rev.* **93**: 957.

Estermann, I., Friedberg, S. A., and Goldman, J. E. (1952).—*Phys. Rev.* **87**: 582.

Gilmont, R. (1951).—*Analyt. Chem.* **23**: 157.

Keesom, W. H., and Kok, J. A. (1936).—*Physica, 's Grav.* **3**: 1035.

Klemens, P. G. (1954).—*Aust. J. Phys.* **7**: 70.

Krutter, H. M. (1935).—*Phys. Rev.* **48**: 664.

Ramanathan, K. G., and Srinivasan, T. M. (1955).—*Phil. Mag.* **46**: 338.

Van Dijk, H., and Schoenberg, D. (1949).—*Nature* **164**: 151.

STRIP INTEGRATION IN RADIO ASTRONOMY

By R. N. Bracewell*

[*Manuscript received January* 17, 1956]

Summary

When a celestial source of radio waves is scanned with an aerial beam which is much longer than the source in one direction but suitably narrow in the other, the transformation from the true distribution to the measured value is referred to as strip integration. It is here treated as a special case of two-dimensional aerial smoothing in which the aerial beam is allowed to spin about its centre as it moves about the sky. It is shown that the resolution obtainable is set by the cross-sectional profile of the strip beam in the narrow dimension. Thus, when the strip reduces to a line, the resolution is complete and full reconstruction of the true distribution is possible; but scans must be made in all directions. In the general case it is shown that there is a principal solution, and that a finite number of scans suffices to determine it. A method is presented for reconstructing the principal solution from the observed data.

I. Introduction

When radio-astronomical observations are made with a fan beam, that is, with one which is long and narrow, the measured quantity is a weighted mean over that part of the celestial brightness distribution lying in the beam. We shall say that the true distribution has been subjected to two-dimensional aerial smoothing by the observational procedure, and it is our purpose to examine the possibilities of recovering the true distribution from the observations in the limiting case where the length of the fan beam greatly exceeds the extent of the source distribution. This special kind of two-dimensional aerial smoothing will be called strip integration. The problem is essentially the same problem of aerial smoothing which has been discussed by Bracewell and Roberts (1954) and by Bracewell (1956) in two papers which will be referred to subsequently as paper I and paper II. However, it is convenient to treat strip integration separately for two reasons. Firstly, strip integration gives not only the blurring effect of averaging over the immediate neighbourhood, but the more serious effect of confusion with distant parts of a source which happen to lie in the long strip. Secondly, when a survey is made with a beam which is circularly symmetrical, no extra information is obtained by rotating the aerial about its main axis. But a fan beam, and particularly a very long fan beam, yields different measurements in different position angles. This property has not yet been utilized in practice with only moderately elongated fan beams, but it is the essence of observations made with the very long fan beams discussed in this paper.

Strip integration has been encountered in a number of recently reported researches, and handled in ways described below, but so far no generally satis-

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney; present address: Radio Propagation Laboratory, Stamford University, Stamford, California.

factory method of inverting the process has been proposed. Wild and Smerd in an interesting theoretical paper presented to the Radio Astronomy Symposium at Jodrell Bank in August 1955, have discussed the relation of a solution, assuming one can be obtained, to the true distribution. Covington and Broten (1954) have used a fan beam 0·12×22·5° lying in the meridian to sweep over the Sun, and have unravelled the strip-integrated solar distribution by an approximate procedure leading to a circularly symmetrical distribution satisfying the observations within the limits of error. They did not deal with the non-symmetrical case nor establish the uniqueness of their solution.

Similar observations by Christiansen and Warburton (1953, 1955) using a fan beam 0·05×5° were carried out so as to sweep over the Sun in different position angles, and they have shown that the distributions along the polar and equatorial diameters are different. They have thus had to face the general problem of no circular symmetry.

The interferometric attempts to get the brightness distribution over the Sun have proceeded similarly from the observations of Stanier (1950) made in a single position angle and reduced on the assumption of circular symmetry, to those of O'Brien (1953) made in many position angles. Machin (1951) advanced the theoretical aspect of Stanier's reduction problem by demonstrating that the desired solution was the Hankel transform of the extrapolated data, and subsequently O'Brien's problem also yielded its solution as the two-dimensional Fourier transform of the extrapolated data (suitably presented). Christiansen and Warburton were able to utilize this work by calculating, by Fourier analysis of their data, what they would have observed had they used a two-aerial interferometer instead of their long array; then the procedure described by O'Brien could be applied. It is a curious fact that this procedure, which removes the effect of strip integration, itself involves further application, numerically, of the same operation (though along a line instead of a strip).

Line integration, which is the limiting case of strip integration as the width of the strip approaches zero, may be expressed as an integral transform as follows :

$$f_L(R,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta+y\sin\theta-R)\mathrm{d}x\mathrm{d}y, \qquad \ldots\ldots \quad (1)$$

where $f_L(R,\theta)$ is the result of line integrating $f(x,y)$ along the line AB (Fig. 1). It is customary to think of $f_L(R,\theta)$ as a function of the continuous variable R for various discrete values of θ, because of the observational method of keeping θ fixed as R varies, that is, of scanning the line AB through positions such as $A'B'$.

The problem now is, given f_L, to find f. If we have f_L for $\theta=\theta_1$ only, an important case because it often arises, the problem is not soluble : for if $f_1(x,y)$ satisfies equation (1) when $\theta=\theta_1$, so also does $f_1(x,y)+f_2(x,y)$, where, using an expression introduced in paper I, $f_2(x,y)$ is an " invisible distribution ", that is, one such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_2(x,y)\delta(x\cos\theta_1+y\sin\theta_1-R)\mathrm{d}x\mathrm{d}y=0. \qquad \ldots\ldots \quad (2)$$

If, however, in addition to $f_L(R,\theta_1)$ we have other information about the source distribution, then further progress may be made. For example, if it is known that the source has circular symmetry then the problem may be fully solved, as explained in Section II. Normally more scans than one will be needed and it has been a matter for discussion how many would suffice to determine a general distribution. Two scans at right angles have been thought enough. Three scans suffice if the source has the kind of elliptical symmetry described by Mills (1953) but as many as six (O'Brien 1953) or eight (Christiansen and Warburton 1953) have been deemed necessary in other cases. It is shown in Section III that scans in *all* directions are needed to solve equation (**1**) for $f(x,y)$. Moreover, given $f_L(R,\theta)$ for all R and θ, $f(x,y)$ is uniquely determined, since there are no invisible distributions, that is, only $f_2(x,y)\equiv 0$ can satisfy equation (**2**) for all θ. However, a finite number of scans suffices in the case of strip integration.

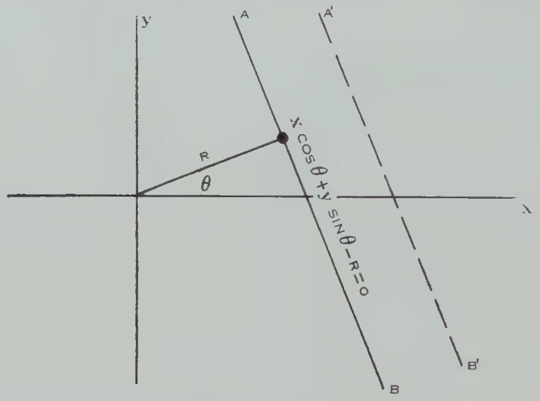


Fig. 1.—Illustrating line integration.

Strip integration, where the line is opened out into a strip with a profile given by the function $A(x)$, is expressed by an integral transform similar to equation (**1**), namely,

$$f_S(R,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)A(x\cos\theta+y\sin\theta-R)dxdy, \quad \ldots . \quad (3)$$

where $f_S(R,\theta)$ is the result of strip integrating $f(x,y)$. The problem of finding $f(x,y)$ given $f_S(R,\theta)$ is more general than the previous one. We expect from paper I to encounter invisible distributions, that is, distributions satisfying

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)A(x\cos\theta+y\sin\theta-R)dxdy=0, \quad \ldots . \quad (4)$$

which will render the solution to equation (**3**) non-unique. However, a " principal solution " can be defined. We also find properties resembling the discrete-interval theorems of papers I and II, which facilitate computing and guide observations.

Among the questions discussed in Section IV are two of current practical importance in reducing observational data: (1) should the scans $f_S(R,\theta_i)$ be restored before proceeding to the solution or should one proceed straight to a solution and consider restoration in two dimensions? (2) Is there a method, resembling the method of successive substitutions, which will allow approximate solutions to be corrected?

II. LINE INTEGRATION WHEN THERE IS CIRCULAR SYMMETRY

In general one is not concerned with sources which show circular symmetry, but, if they do, then one scan leads to a result; and often only one scan is available. This has been so in the early stages of several investigations in radio astronomy, for example, those of Covington, Stanier, and Christiansen and Warburton. Furthermore the same equation arises in another connexion in radio astronomy where a distant spherically symmetrical distribution is integrated along lines of sight (Bolton and Westfold 1951). Other reasons for studying this case before proceeding to the general problem, are that the numerical method of solution proposed here (1) leads to the method later proposed for the general problem and (2) is itself needed in dealing with the symmetrical component of a general distribution.

Let the brightness temperature distribution $f(x,y)$ depend only on r, where $r^2=x^2+y^2$, and let it be written $f(r)$. Then the line integrated distribution $f_L(R,\theta)$, given by equation (**1**), will be the same for all θ. Put $\theta=0$ in equation (**1**) and write $f_L(x)$ for $f_L(R,0)$. Then

$$f_L(x)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(r)\delta(x-R)\mathrm{d}x\mathrm{d}y,$$

$$=2\int_{r=x}^{\infty} f(r)\mathrm{d}y,$$

or

$$f_L(x)=2\int_x^{\infty}\frac{f(r)r\mathrm{d}r}{(r^2-x^2)^{\frac{1}{2}}}. \qquad (\mathbf{5})$$

This is a form of Abel's integral equation and for convenience we may refer to $f_L(x)$ as the Abel transform of $f(r)$. An explicit solution for $f(r)$ may be found by converting the integral into a convolution integral. Put $\xi=x^2$ and $\rho=r^2$, and let $f_L(x)=F_L(x^2)$, $f(r)=F(r^2)$. Then equation (**5**) is replaced by

$$F_L(\xi)=\int_{-\infty}^{\infty} K(\xi-\rho)F(\rho)\mathrm{d}\rho,$$

or

$$F_L=K*F, \qquad (\mathbf{6})$$

where

$$K(\xi)=\begin{cases}(-\xi)^{-\frac{1}{2}} & (\xi<0),\\ 0 & (\xi\geqslant 0).\end{cases}$$

Equation (6) may be solved by taking the Laplace transform of both sides, solving the resulting algebraic equation for the transform of F, and retransforming. Then

$$F=-\frac{1}{\pi}K*F_L', \qquad \text{(7)}$$

or, alternatively, by a different choice of the factors of the transform of F,

$$F=\frac{2}{\pi}K^{-1}*F_L'', \qquad \text{(8)}$$

a solution which may also be obtained directly from equation (7) by integrating by parts. The solution is unique and there are therefore no invisible distributions (other than null functions). Reverting to f and f_L, we may write equations (7) and (8) as

$$f(r)=-\frac{1}{\pi}\int_r^{\infty}\frac{f_L'(x)\mathrm{d}x}{(x^2-r^2)^{\frac{1}{2}}}=-\frac{1}{\pi}\int_r^{\infty}(x^2-r^2)^{\frac{1}{2}}\frac{\mathrm{d}}{\mathrm{d}x}\left\{\frac{f_L'(x)}{x}\right\}\mathrm{d}x,$$

or, if the integrand is zero beyond $x=r_0$, and allowing for the possibility that the integrand may behave impulsively at r_0,

$$f(r)=-\frac{1}{\pi}\int_r^{r_0}\frac{f_L'(x)\mathrm{d}x}{(x^2-r^2)^{\frac{1}{2}}}+\frac{f_L(r_0)}{\pi(r_0^2-r^2)^{\frac{1}{2}}},$$

$$=\frac{1}{\pi}\int_r^{r_0}(x^2-r^2)^{\frac{1}{2}}\frac{\mathrm{d}}{\mathrm{d}x}\left\{\frac{f_L'(x)}{x}\right\}\mathrm{d}x-\frac{f_L'(r_0)}{\pi r_0}(r_0^2-r^2)^{\frac{1}{2}}.$$

The following table of Abel transforms was worked out from the formulae. In the first eight examples f and f_L are zero for r and x greater than a.

These results are illustrated in Figure 2 which also has graphs of F and F_L. A useful relation for checking Abel transforms is

$$\int_{-\infty}^{\infty}f_L(x)\mathrm{d}x=2\pi\int_0^{\infty}f(r)r\mathrm{d}r;$$

also

$$f_L(0)=2\int_0^{\infty}f(r)\mathrm{d}r.$$

Another property is that

$$K*K*F'=-\pi F, \qquad \text{(9)}$$

that is, the operation $K*$ applied twice in succession annuls differentiation; thus $K*$ is, in this sense, half-order integration. Conversely, F is the half-order differential coefficient of F_L. Fractional order differentiation has been extensively studied since a publication by Laplace in 1812 (e.g. see Doetsch 1943, p. 298), and as a result of its application to electric circuit problems by Heaviside (1922), is well known in modern circuit analysis (see e.g. Bush 1937 ; Starr 1953).

In our notation equation (9) is easily derived, for if $F_L = K*F$ implies $F = -\pi^{-1}K*F_L'$, then it follows further that $F_L' = K*F'$; whence

$$K*K*F' = K*F_L' = -\pi F.$$

When it becomes necessary to perform line integration numerically, advantage may be taken of the possibility of expressing equation (5) in terms of the convolution $K*F$. Approximate evaluation of $K*F$ by taking sums of products of $K(\rho)$ and $K(\xi-\rho)$ at discrete intervals of ρ is then facilitated by the fact that the numerical values of K are the same for all ξ. Furthermore the values are the same, however fine an interval is chosen, save for a normalizing factor. Consequently a universal table of values can be set up for permanent reference. Table 2 shows coefficients for immediate use with values of F read off at $\rho = \frac{1}{2}, 1\frac{1}{2}, \ldots, 9\frac{1}{2}$, the scale of ρ being such that F becomes zero or negligible at $\rho = 10$. The table gives mean values of K over the intervals 0–1, 1–2, . . . Thus at $\rho = n + \frac{1}{2}$ the value is

TABLE 1

ABEL TRANSFORMS

	$f(r)$		$f_L(x)$
Disk	1	Semi-ellipse	$2(a^2-x^2)^{\frac{1}{2}}$
	$(a^2-r^2)^{-\frac{1}{2}}$	Rectangle	π
Hemisphere	$(a^2-r^2)^{\frac{1}{2}}$	Parabola	$\frac{1}{2}\pi(a^2-x^2)$
Paraboloid	a^2-r^2		$\frac{4}{3}(a^2-x^2)^{3/2}$
	$(a^2-r^2)^{3/2}$		$\frac{3\pi}{8}(a^2-x^2)^2$
Cone	$a-r$		$a(a^2-x^2)^{\frac{1}{2}}-x^2\cosh^{-1}\frac{a}{x}$
	$\frac{1}{\pi}\cosh^{-1}\frac{a}{r}$	Triangle	$a-x$
	$\delta(r-a)$		$2a(a^2-x^2)^{-\frac{1}{2}}$
Gaussian	e^{-r^2/σ^2}	Gaussian	$\sigma\pi^{\frac{1}{2}}e^{-x^2/\sigma^2}$
	$r^2e^{-r^2/\sigma^2}$		$\sigma\pi^{\frac{1}{2}}(x^2+\frac{1}{2}\sigma^2)e^{-x^2/\sigma^2}$
	$\frac{r^2-\frac{1}{2}\sigma^2}{\sigma\pi^{\frac{1}{2}}}e^{-r^2/\sigma^2}$		$x^2e^{-x^2/\sigma^2}$
	$(b^2+r^2)^{-1}$		$\pi(b^2+x^2)^{-\frac{1}{2}}$
	$J_0(\omega r)$		$\frac{2}{\omega}\cos\omega x$
	$M\left(\frac{\omega r}{\pi}\right)$		$\frac{\sin^2\omega x}{\omega^2x^2}$

$$\int_n^{n+1} K(-\rho)d\rho = 2(n+1)^{\frac{1}{2}} - 2n^{\frac{1}{2}}.$$

Where N points of subdivision are used, the scale of ρ is arranged so that F becomes zero at $\rho = N$ and the coefficients are all multiplied by $(10/N)^{\frac{1}{2}}$.

As an example consider $F(\rho)=(10-\rho)^{\frac{1}{2}}$, for which the solution is known to be $F_L(\xi)=\frac{1}{2}\pi(10-\xi)$. We work at unit intervals of ρ, and copy the coefficients on a movable strip. The calculation in progress is shown in Table 3. The movable strip is in position for calculating $F_L(\xi)$ as the sum of products of corresponding values of F and K $(7\cdot78=2\cdot12\times2\cdot000+1\cdot87\times0\cdot828+\ldots+0\cdot71\times0\cdot472)$. This method for calculating Abel transforms numerically is very quick.

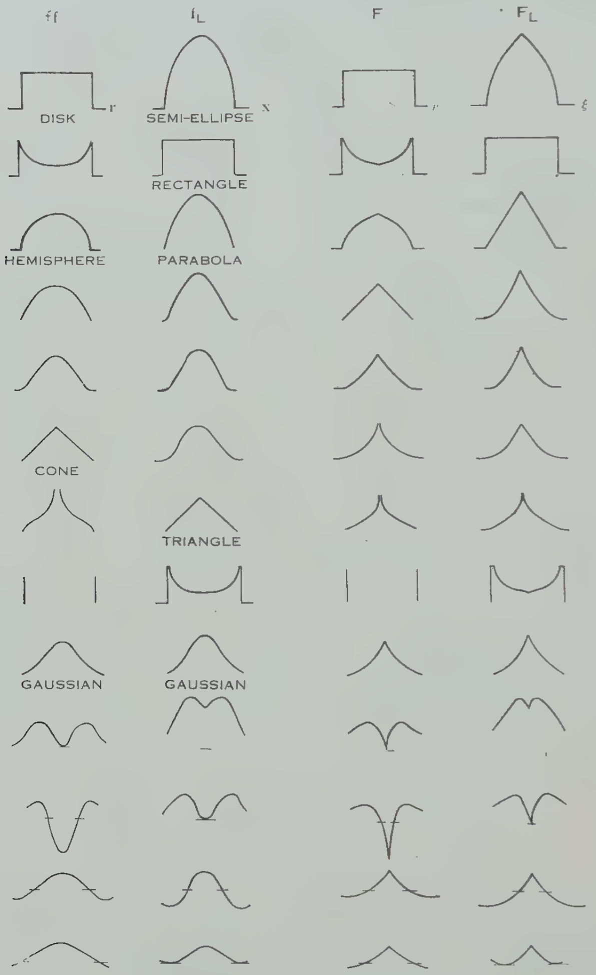


Fig. 2.—Some Abel transforms.

To perform the inverse operation, however, is the problem facing us; i.e. given F_L, to find F. We have equation (7) which leads directly to a solution if F_L is first differentiated. But if Table 3 is studied it will be perceived that the calculation just described can be done in reverse, using the values of F_L and working the movable strip upwards from the bottom. The strip is shown in position for calculating $F(5\frac{1}{2})$, let us say on a hand calculating machine. Form the products $0\cdot71\times0\cdot472, \ldots, 1\cdot87\times0\cdot828$, allowing them to accumulate in the product register. Clear the multiplier register, set $2\cdot000$ on the setting levers, and turn until $7\cdot78$ shows in the product register. The value $F(5\frac{1}{2})=2\cdot12$ then shows in the multiplier register. The inverse transformation performed in this way is practically as quick as the direct transformation.

TABLE 2

COEFFICIENTS FOR PERFORMING OR INVERTING THE ABEL TRANSFORMATION

ρ	K	ρ	K	ρ	K	ρ	K
$\frac{1}{2}$	2·000	$5\frac{1}{2}$	0·427	$10\frac{1}{2}$	0·309	$15\frac{1}{2}$	0·254
$1\frac{1}{2}$	0·828	$6\frac{1}{2}$	0·393	$11\frac{1}{2}$	0·295	$16\frac{1}{2}$	0·246
$2\frac{1}{2}$	0·636	$7\frac{1}{2}$	0·364	$12\frac{1}{2}$	0·283	$17\frac{1}{2}$	0·239
$3\frac{1}{2}$	0·536	$8\frac{1}{2}$	0·343	$13\frac{1}{2}$	0·272	$18\frac{1}{2}$	0·233
$4\frac{1}{2}$	0·472	$9\frac{1}{2}$	0·325	$14\frac{1}{2}$	0·263	$19\frac{1}{2}$	0·226

TABLE 3

CALCULATING ABEL TRANSFORMS

ρ	F	K	F_L
			15·65
	3·08		
1			14·08
	2·91		
2			12·52
	2·74		
3			10·94
	2·55		
4			9·37
	2·35		
5		$\longrightarrow$	7·78
	2·12	2·000	
6			6·20
	1·87	0·828	
7			4·62
	1·58	0·636	
8			3·03
	1·22	0·536	
9			1·42
	0·71	0·472	
10			0
		0·427	
		·········	

It is hardly necessary to point out that $f_L(x)$ depends only on values of $f(r)$ for which $r \geqslant x$. When we are given $f_L(x)$ for all x and we know $f(r)$ over an outer annulus (shown shaded in Fig. 3) then the value of $f(r)$ just inside the known region can be deduced from the integral along AB. This in effect is what the calculation does, and the graphical interpretation of it suggests a way of handling the problem when there is no circular symmetry.

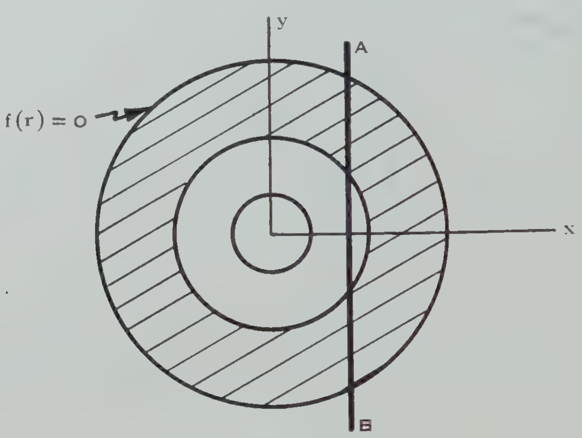


Fig. 3.—Contour diagram of $f(r)$ in the xy-plane.

III. Line Integration in the Absence of Symmetry

In this section we deal frequently with the following transforms which are therefore defined here for reference.

Fourier transform : $$\bar{f}(s)=\int_{-\infty}^{\infty} e^{i2\pi sx}f(x)\mathrm{d}x, \qquad \bar{f}=\mathbf{F}f. \quad \ldots (10)$$

Two-dimensional Fourier transform : $$\bar{f}(u,v)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{i2\pi(ux+vy)}\mathrm{d}x\mathrm{d}y, \quad \bar{f}={}^2\mathbf{F}f. \quad \ldots (11)$$

Hankel transform of zero order : $$f_H(\rho)=2\pi\int_0^{\infty} f(r)r\mathrm{J}_0(2\pi\rho r)\mathrm{d}r, \qquad f_H=\mathbf{H}f. \quad \ldots (12)$$

Abel transform : $$f_L(s)=2\int_s^{\infty}\frac{f(x)x\mathrm{d}x}{(x^2-s^2)^{\frac{1}{2}}}, \qquad f_L=\mathbf{A}f. \quad \ldots (13)$$

A great gain in clarity can be had in some places by the use of the abbreviated notation in terms of functional operators, shown on the right. The Abel transform has been discussed in Section II, but a note on the two-dimensional Fourier transform, the Hankel transform, and their relationship might be useful. Two-dimensional Fourier analysis decomposes a function of x and y into wave components such as that shown in Figure 4. Each such component is represented in amplitude and phase by a complex number on the uv-plane at a point whose polar coordinates (R,θ) represent respectively the wave number and the direction of the wave normal. When the function of x and y is circularly symmetrical, so also will its two-dimensional Fourier transform be, that is, both the function and its transform are functions of one (radial) variable only. On expressing this fact in equation (**11**) and integrating over the angular coordinate we find that the relationship between the two functions of radius only is equation (**12**). Thus if two circularly symmetrical functions are two-dimensional Fourier transforms, one of the other, then their cross-sectional profiles are Hankel transforms, one of the other.

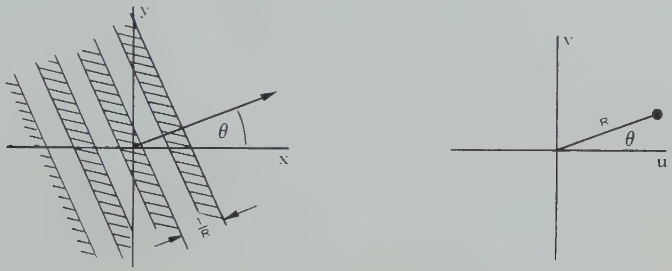


Fig. 4.—A two-dimensional Fourier component and its representation on the *uv*-plane.

The Hankel transformation is reciprocal, that is, $f_H=\mathbf{H}f$ and $f=\mathbf{H}f_H$, or $\mathbf{HH}f=f$, or

$$\mathbf{HH}=\mathbf{I},$$

where **I** is the operator which leaves a function unchanged. The operators **F** and $^2\mathbf{F}$ are not quite reciprocal but we may nevertheless write

$$\mathbf{FF}={}^2\mathbf{F}\,{}^2\mathbf{F}=\mathbf{I},$$

if the i in equations (**10**) and (**11**) is changed to $-$i before repeating the operation. The operator **A** is not reciprocal and if applied twice is equivalent to integration (with respect to the proper variable as follows from Section II):

$$-\pi^{-1}\frac{\mathrm{d}}{\mathrm{d}(x^2)}\mathbf{AA}=\mathbf{I}.$$

Having noted these properties of the various transforms we can proceed to express line integration in terms of them. Let $f(x,y)$ be a true distribution of brightness temperature, as indicated roughly in Figure 5 by contours. The line integral $f_L(R,\theta)$ is given by equation (**1**) in which we shall put $\theta=0$. Let $f_L(x)=f_L(R,0)$. Then $f_L(x)$, also shown in Figure 5, is the result of line integrating $f(x,y)$ along lines $x=$constant. Thus

$$f_L(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y.$$

Consider now the Fourier transform of $f_L(x)$.

$$\begin{aligned}\mathbf{F}f_L&=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}2\pi ux}\mathrm{d}x\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}2\pi ux}f(x,y)\mathrm{d}x\mathrm{d}y\\&=\bar{f}(u,0),\end{aligned}$$

where $\bar{f}(u,v)$ is ${}^2\mathbf{F}f$. If now the contour diagram in the lower right-hand corner of Figure 5 stands for $\bar{f}(u,v)$, then $\bar{f}(u,0)$, the Fourier transform of $f_L(x)$, is the cross section along $v=0$. This cross section is shown rabatted into the uv-plane and cross hatched.

The result of line integrating at a different value of θ, for example along lines parallel to OP, is shown at the left on tilted axes. The Fourier transform of this curve is the cross section of $\bar{f}(u,v)$ along QR. We have now demonstrated the relationship of line integration to the Fourier transformation, and we can use this relationship to reach some immediate conclusions.

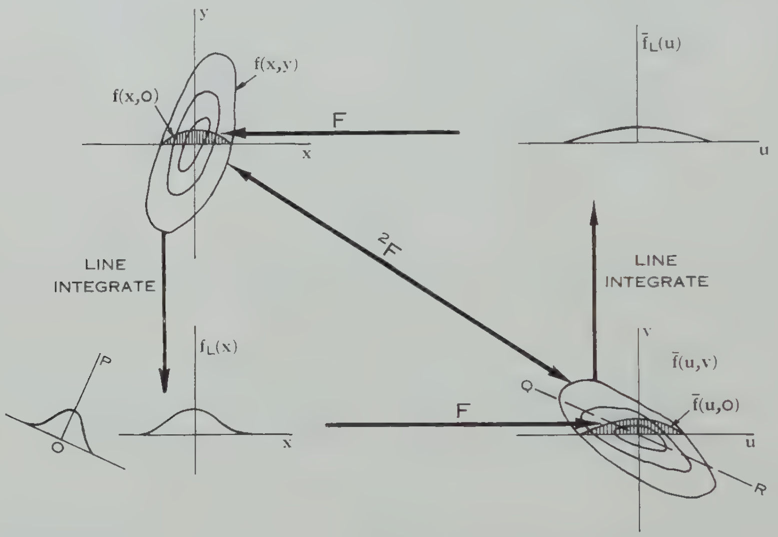


Fig. 5.—Synoptic chart of the relationship of line integration to one- and two-dimensional Fourier transforms.

Because of the uniqueness with which a function is defined by its Fourier transform it is clear that line integration for *all* θ is necessary to determine *every* cross section of $\bar{f}(u,v)$, without which $f(x,y)$ will not be determined. Hence all scanning directions must be used in order to find $f(x,y)$ from $f_L(R,\theta)$. Furthermore, when this is done, $f(x,y)$ is determined uniquely. Thus there can be no invisible distributions, in the non-symmetrical as in the symmetrical case.

Studies made with a two-aerial interferometer, whose beam pattern on the xy-plane is as in Figure 4, start with measurements of $\bar{f}(u,v)$ as the basic data. To find $f(x,y)$ it is necessary to take the two-dimensional Fourier transform. This one can do by line integrating and taking one-dimensional Fourier transforms, that is, by taking the path from $\bar{f}$ to f via $\bar{f}_L$ in Figure 5. Alternatively one might take Fourier transforms and perform the operation inverse to line integration, that is, proceed via f_L. The former procedure was introduced into radio astronomy by O'Brien (1953). By assembling all the different steps into the one diagram we obtain a convenient summary of the situation.

When $f(x,y)$ is circularly symmetrical we find further connexions among the operations, which are applicable not only to symmetrical sources but to symmetrical aerial patterns and to observational procedures which are symmetrical.

In the first place, line integration becomes Abel transformation. Then the cross sections of $f(x,y)$ and $\bar{f}(u,v)$, being independent of θ, become Hankel transforms. Figure 6 summarizes this situation. The basic theorem brought out by this discussion is*

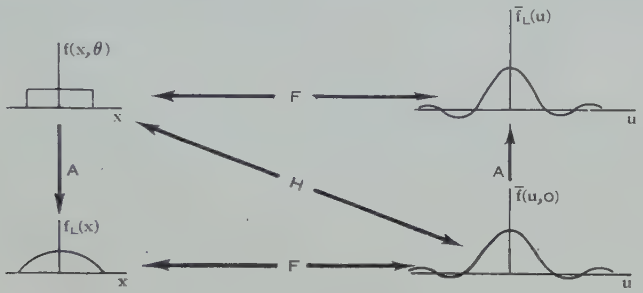


Fig. 6.—Relationships between the Fourier, Abel, and Hankel transformations.

$$\mathbf{HFA}=\mathbf{I},$$

or

$$\int_0^\infty \mathrm{d}r J_0(2\pi\xi r)r\int_{-\infty}^\infty \mathrm{d}s e^{i2\pi rs}\int_s^\infty \frac{\mathrm{d}x 2xf(x)}{(x^2-s^2)^{\frac{1}{2}}}=f(\xi).$$

Several other corollaries can be written down at sight.

$$\mathbf{HFA}=\mathbf{FAH}=\mathbf{AHF}=\mathbf{I},$$
$$\mathbf{H}=\mathbf{FA}\quad \mathbf{F}=\mathbf{AH}\quad \mathbf{A}=\mathbf{FH},$$
$$\mathbf{AFAF}=\mathbf{FAFA}=\mathbf{I},$$
$$\mathbf{AFA}=\mathbf{F},\quad \mathbf{A}=\mathbf{FA}^{-1}\mathbf{F}.$$

Some of these may be expressed in words as follows. The Fourier transform of the Abel transform is the Hankel transform. Abel transformation in the function domain corresponds to inverse Abel transformation in the Fourier transform domain. Table 4 gives four cyclical sets of functions related as in Figure 6, which itself illustrates the first set of functions in the table. The functions have been chosen because of their tendency to occur in the present work; thus the table summarizes four Hankel, seven Fourier, and eight Abel transform pairs.

How to perform the operation inverse to line integration, which may be referred to as reconstruction, may now be considered. Clearly one way of proceeding from f_L to f is via $\bar{f}$ and $\bar{f}_L$. Figure 7 summarizes what one must do, namely, (i) take Fourier transforms of all the line integrated profiles f_L,

---

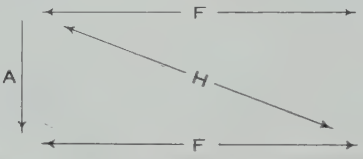
* The compound operator **HFA** has the effect of applying the operations **A**, **F**, **H** in that order, i.e. $\mathbf{HFA}f=\mathbf{H}[\mathbf{F}(\mathbf{A}f)]$.

(ii) assemble the transforms into a two-dimensional function $\bar{f}$, (iii) line integrate $\bar{f}$ for all directions, (iv) take Fourier transforms of all the line-integrated profiles $\bar{f}_L$, and (v) assemble the transforms into the two-dimensional solution f. We shall refer to this as reconstruction by the roundabout path.

Direct reconstruction can be effected, in the case of circular symmetry, by a numerical method which is equivalent to the graphical method of Figure 3, and by a simple extension this method can be made applicable in the absence of symmetry.

TABLE 4

FOURIER-ABEL-HANKEL SETS*



$\Pi\left(\frac{x}{2F}\right)$	$\frac{\sin 2\pi Fu}{\pi u}$
$2(F^2-x^2)^{\frac{1}{2}}\Pi\left(\frac{x}{2F}\right)$	$\frac{FJ_1(2\pi Fu)}{u}$
$\frac{\sin 2\pi Fx}{\pi x}$	$\Pi\left(\frac{u}{2F}\right)$
$J_0(2\pi Fx)$	$\frac{1}{\pi(F^2-u^2)^{\frac{1}{2}}}$
$\frac{1}{2}\delta(x-F)$	$\cos 2\pi Fu$
$\frac{F}{(F^2-x^2)^{\frac{1}{2}}}$	$\pi FJ_0(2\pi Fu)$
$M(Fx)$	$F(F^2-u^2)^{\frac{1}{2}}-u^2\cosh^{-1}\frac{F}{u}$
$\frac{\sin^2 \pi Fx}{\pi^2 x^2}$	$F\Lambda\left(\frac{u}{2F}\right)$

* In this table $\Pi(x)$ and $\Lambda(x)$ are respectively the rectangle and triangle functions of unit height and width, and $M(x)$ is defined in Section IV.

Let f be zero (or negligibly small) outside a closed contour C. Then starting from C one could work inwards proceeding at each step as explained in connexion with Figure 3. But C is not known; instead one knows that $f_L(R,\theta)$ is zero outside a curve L which may be seen to be the locus of the foot of the perpendicular from the origin on to those tangents to C which do not cut C. Given L one cannot determine C, but C will lie within, and largely coincide with, the envelope E of lines through the points of L perpendicular to the lines joining these points to the origin. The curve L is pedal to E, a relationship which was first pointed out in connexion with line integration by Dr. R. F. Mullaly.

Since points of C can lie outside neither E nor L, either of the latter may be made the starting boundary for reconstruction. Some points of technique facilitate the process. For example, it will often happen that there is approximate circular symmetry over a sector. Then in that sector the numerical procedure described earlier may be applied for some distance in. This method applies especially at the outer boundary where it may be used to get the process started. A coarse interval is suggested so that an approximate result can be obtained with the minimum computation ; then a test akin to that used in the method of successive substitutions can be applied to determine corrections. We now pause to consider application of the method of successive substitutions.

If φ is an approximation to f, then φ_L, which is obtained by line integrating φ, will be an approximation to f_L. Apply $f_L-\varphi_L$, suitably normalized, as a correction to φ at (R,θ) to give φ_1 as a new approximation which may then in its turn be corrected. Then the question is, do the successive approximations φ, φ_1, etc. tend to the solution f? It can be shown that even for the case of circular symmetry this is not in general so, and a modification has to be sought. A satisfactory outcome results if one incorporates the correction at each point before proceeding to correct the next, and if one works inwards in such a manner that at each step the value to be corrected is the only uncorrected value on the line of integration. The φ_1 so obtained must then be equal to f (to the order of accuracy permitted by the coarseness of the intervals).

$$\begin{tikzcd}
f \arrow[d, "\text{LINE INTEGRATE}"'] & & \bar{f}_L \arrow[ll, "F"'] \\
f_L \arrow[rr, "F"'] & & \bar{f} \arrow[u, "\text{LINE INTEGRATE}"']
\end{tikzcd}$$

Fig. 7.—The inversion of line integration.

Considerable work is involved in performing direct reconstruction as described above, but comparison with the only other known method is favourable; for reconstruction by the roundabout path involves precisely the same amount of line integration together with a large number of Fourier transformations and graphical manipulations. Attention to the detail of performing line integration is indicated, and in the next section it will be shown how line integration in radio astronomy can in practice be reduced exactly to a summation.

IV. STRIP INTEGRATION

A diagram is useful for surveying the complexities of strip integration. It is clear, even without reference to equations (**1**) and (**3**), that the strip-integrated function $f_S(R,\theta)$ is related to $f_L(R,\theta)$ by convolution with the profile A of the aerial beam. Thus

$$f_S(R,\theta)=A(R)*f_L(R,\theta).$$

We may put $\theta=0$ and write $f_S(x)$ for $f_S(R,0)$. Figure 8 shows $f_S(x)$ on an extended diagram of the type of Figure 5. The Fourier transforms of $f_L(x)$ and $f_S(x)$, which are shown to the right, are related through multiplication by $\bar{A}$.

The profiles $f_S(x)$ are observed for different values of θ and it is then required to find $f(x,y)$.

Method I.—One may first restore the profiles $f_S(x)$ as in paper I or, preferably, by the chord construction (Bracewell 1955), and then reconstruct as in Section III.

Method II.—Alternatively, one might reconstruct first and then restore as in paper II, but the restoration would be two-dimensional and not as convenient. However, as the value of restoration is sometimes doubtful, it may in some cases be desirable to postpone considering restoration until after reconstruction. The question then arises how such restoration may be done.

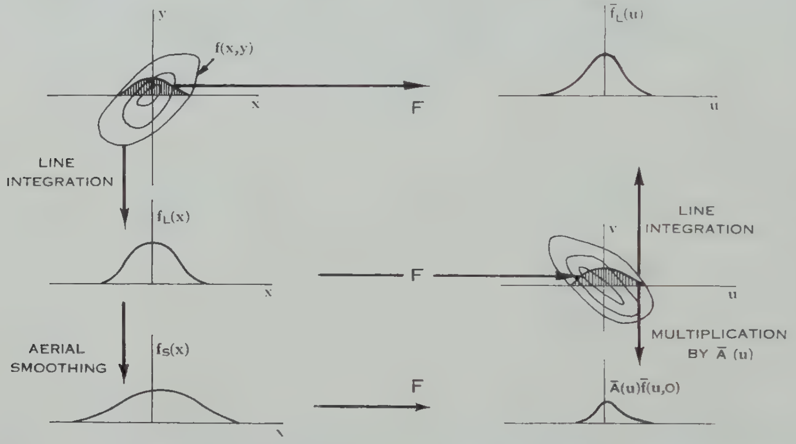


Fig. 8.—Strip integration considered as line integration followed by aerial smoothing.

Method III.—The only method so far applied has been as follows: (i) take Fourier transforms of all profiles $f_S(x)$, (ii) correct each transform for aerial smoothing by dividing by $\bar{A}$, (iii) assemble into a two-dimensional contour diagram, (iv) line integrate in all directions, (v) take Fourier transforms of all line-integrated profiles, (vi) assemble the transforms into a contour diagram representing the solution. As before, one might defer restoration until the end. The labour involved in this roundabout method is so great that it will probably fall into disuse now that a direct procedure is available.

The interesting question raised regarding the commutativity of restoration and reconstruction may now be studied. If a line-integrated profile $f_L(x)$ had a Fourier transform which cut out to zero at $u=u_c$, then the two-dimensional Fourier transform of $f(x,y)$ would fall to zero at $u=u_c$, $v=0$ (see Fig. 5). Now $f_S(x)$, which is derived from $f_L(x)$ by aerial smoothing, must have this property, and similarly for the strip-integrated profiles at all values of θ. The cut-off value of $\sqrt{(u^2+v^2)}$ will be the same for all θ; hence the distribution over the uv-plane, of which $\bar{A}(u)\bar{f}(u,0)$ in Figure 8 is the $\theta=0$ cross section, must have a circular boundary and be zero outside. Call this function $\overline{F}(u,v)$. Let the aerial pattern be given by

$$A(x)=\left(\frac{\sin \pi x u_c}{\pi x u_c}\right)^2.$$

Then

$$\bar{A}(u)=\Lambda\left(\frac{u}{2u_c}\right),$$

where $\Lambda(u)$ is the triangle function of unit height and base, that is,

$$\Lambda(u)=\begin{cases}1-2\,|\,u\,| & (|\,u\,|<\tfrac{1}{2})\\ 0 & (|\,u\,|>\tfrac{1}{2}).\end{cases}$$

Thus

$$\overline{F}(u,v)=\Lambda\left\{\frac{\sqrt{(u^2+v^2)}}{2u_c}\right\}\bar{f}(u,v).$$

If now one proceeds, without restoration, to reconstruct from $f_S(x)$, the result will be $F(x,y)$ instead of $f(x,y)$, where $F(x,y)$ is the two-dimensional Fourier transform of $\overline{F}(u,v)$. By the two-dimensional convolution theorem

$$F=M*f,$$

where $M(Fr)$ is the Hankel transform of $\Lambda(\rho/2F)$:

$$M=\mathbf{H}\Lambda.$$

Consequently F differs from f through being smoothed with a pattern $M(Fr)$; hence f can be recovered as well by restoration following reconstruction as otherwise. This fact was demonstrated by Wild and Smerd.

From Table 4 we see that M may easily be calculated numerically as the inverse Abel transform of $\pi^{-2}x^{-2}\sin^2\pi F x$, but the Hankel transform may in this case also be readily performed in terms of tabulated functions. Thus

$$\begin{aligned}M(Fr)&=2\pi\int_0^\infty \rho\Lambda\left(\frac{\rho}{2F}\right)\mathrm{J}_0(\rho r)\mathrm{d}\rho\\ &=2\pi\int_0^F \rho\left(1-\frac{\rho}{F}\right)\mathrm{J}_0(\rho r)\mathrm{d}\rho.\end{aligned}$$

After integration by parts, and with some reduction,

$$M(\zeta)=2\pi F^2\{\zeta^{-3}j(\zeta)-\zeta^{-2}\mathrm{J}_0(\zeta)\},$$

where $\zeta=Fr$ and

$$j(\zeta)=\int_0^\zeta \mathrm{J}_0(\zeta)\mathrm{d}\zeta,$$

a tabulated function.*

---

* See, for example, Watson (1944) or National Bureau of Standards (1954).

All the functions and operations involved in the above discussion can be summarized for reference by arranging them on the vertices and edges of a cube as in Figure 9 (*a*). Using the particular aerial pattern assumed above, we illustrate in Figure 9 (*b*), the four equivalent functions representative of the aerial, and the relations between them. This figure is taken from the last set of functions of Table 4. Figure 9 (*c*) shows how methods I, II, and III described above may be stated diagrammatically as paths on the cube. O'Brien's method is also indicated. On this diagram it is easy to see alternative paths from one point to another, and to study their significance. Some of the less obvious relationships so revealed might well prove useful in future deliberations.

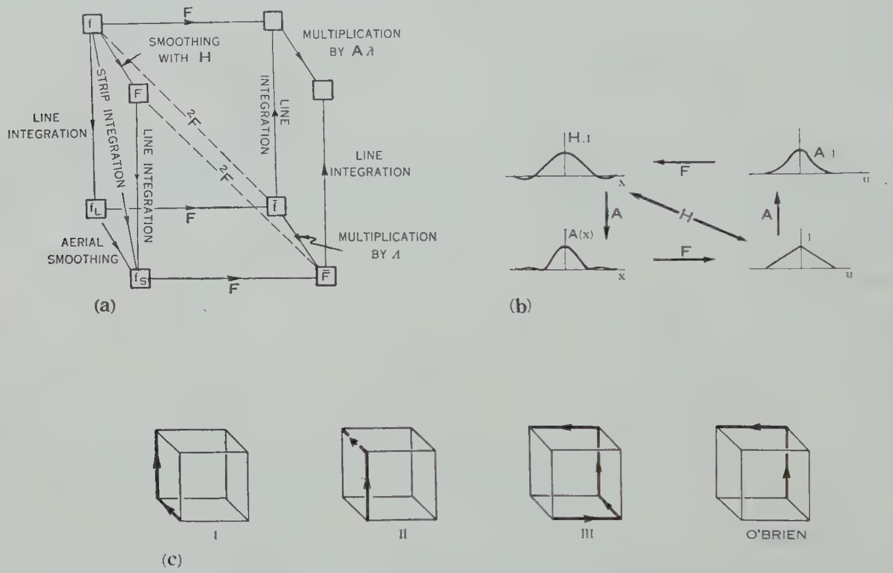


Fig. 9.—Relationships between functions involved in strip integration.

In Figure 9 (*a*) all four operations represented by lines parallel to that joining f_L to f_S are irreversible, since information is lost in the process. Therefore when in Figure 9 (*c*) we show arrows running in the reverse direction, for example from F to f, it is understood that one will not recover f itself but an approximation to f whose Fourier transform is zero outside a certain region.

We have seen how in principle to invert strip integration, and now prove a theorem which adds greatly to the practicability of the method.

If a function f(x) *has a Fourier transform* $\bar{\mathrm{f}}(\mathrm{s})$ *which is zero for* $\mathrm{s}>\tau_0^{-1}$, *then* $\int_{-\infty}^{\infty}\mathrm{f}(\mathrm{x})dx$ *is equal to* τ *times the sum of values of* f(x) *separated by interval* τ, *where* τ *is any interval shorter than* τ_0. For

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x=\bar{f}(0)\,;$$

but

$$\tau \mathrm{III}(\tau s) * \bar{f}(s) = \bar{f}(s), \qquad s=0,$$

provided $\tau < \tau_0$. Therefore, since the Fourier transform of $\tau \mathrm{III}(\tau s) * \bar{f}(s)$ is $\mathrm{III}(x/\tau) f(x)$,

$$\int_{-\infty}^{\infty} f(x) \mathrm{d}x = \int_{-\infty}^{\infty} \mathrm{III}\left(\frac{x}{\tau}\right) f(x) \mathrm{d}x = \tau \sum_{n=-\infty}^{\infty} f(x - n\tau).$$

As in papers I and II, the row-of-spikes function $\mathrm{III}(x)$ is defined by

$$\mathrm{III}(x) = \sum_{n=-\infty}^{\infty} \delta(x+n).$$

The interval τ_0 is twice the peculiar interval at which discrete values of $f(x)$ suffice to specify $f(x)$ and is therefore quite coarse: for $x^{-2} \sin^2 x$ we have $\tau_0 = \pi$ and for $x^{-1} \sin x$, $\tau_0 = 2\pi$. If one takes $\tau = \tau_0$ then the summation will be subject to a correcting factor $(1+D)^{-1}$ if there is an average discontinuous jump $D\bar{f}(0)$ in $\bar{f}(s)$ at $s = \pm \tau_0^{-1}$; in the case of $x^{-1} \sin x$, $D=1$.

In line integrating a function $\varphi(x,y)$ in the course of reconstruction by method I, the integral is correctly evaluated by summing discrete values taken along the straight line at intervals which are equal to twice the peculiar interval of the aerial as defined in paper I. This follows from the theorem because in method I the line integrals are compared with points on strip-integrated profiles which are free from Fourier components of semi-period greater than the peculiar interval of the aerial.

The possibility of obtaining line integrals correctly by quite coarse summing is directly helpful, but has another implication which is no less important. Since only discrete information is utilized, scans in all directions are not required; only enough scans are wanted to permit the process of reconstruction. On this basis one can answer the question of how many position angles should be programmed for observation.

Let the units of x and y equal the peculiar interval associated with A and let the major diameter of the source be D. If N position angles, equally spaced between 0 and π are used, then, at the boundary of the source, observations will be spaced by at most $\frac{1}{2}\pi D/N$. This spacing must not exceed the spacing needed for the summation which gives the line integral. Hence

$$\pi D/2N \leqslant 2, \text{ or } N \geqslant \pi D/4.$$

For example, if the diameter of the Sun is 20 peculiar intervals, at least 16 different position angles are necessary. In the case of symmetrical objects like the quiet Sun, the number of necessary directions is halved.

A final remark bearing on the practical details of inverting strip integration may be added. Since the Abel transformation may be so readily inverted, by the method of Section II, a symmetrical component may be worth subtracting from the data. If the unsymmetrical residual is small in comparison, then not such high accuracy is required in inverting the strip integration as when the whole is handled at once. In the author's opinion this method would be suitable in the case of the Sun.

V. FAN BEAMS

The typical fan beam is that produced by a rectangular array or rectangular aperture, and if the aperture distribution is uniform the aerial pattern is approximately of the form

$$\frac{\sin k_1x}{x} \cdot \frac{\sin k_2y}{y}.$$

When such a beam is displaced parallel to itself, the situation can be reduced approximately to that of a circular beam by working in terms of the peculiar interval appropriate to each of the x and y directions. But when the beam may rotate about its centre a new factor is introduced. More information is available since there are now three independent variables (x, y, and ψ) instead of two (x and y). The more elongated the beam the more the additional information, but on the other hand, the elongation itself means a reduction in information. It will now be shown that the gain and loss counterbalance each other exactly.

Consider a circularly symmetrical beam $A\{\sqrt{(x^2+y^2)}\}$ with Hankel transform $B\{\sqrt{(u^2+v^2)}\}$ which falls to zero where $\sqrt{(u^2+v^2)}=P$. From a survey of a distribution $f(x,y)$ with such a beam, $\bar{f}(u,v)$ can be recovered out to a distance P from $u=0$, $v=0$. Now let the beam be elongated by a factor b in the y direction. The new beam $A\{\sqrt{(x^2+b^{-2}y^2)}\}$ has a Hankel transform proportional to $B\{\sqrt{(u^2+b^2v^2)}\}$ and from a survey with such a beam less of $\bar{f}(u,v)$ is recoverable, but it is still recoverable out to a distance P in one direction. As the elongated beam is turned through all ψ, $\bar{f}(u,r)$ is recovered out to a distance P in all directions on the uv-plane. Hence as much information about $f(x,y)$ is obtainable by spinning the elongated beam as is yielded by the small circular beam.

If the elongated beam is produced by a rectangular array then it does not reduce exactly to a circularly symmetrical beam when contracted, so the situation is a little more complicated. It can be shown that the information obtained by spinning such a beam exceeds what is obtained with the small non-elongated beam which would be produced by stretching the rectangular array out into a square. But the difference is exactly compensated by the extra information available from the non-elongated beam itself, if it is spun.

Two special cases of this fan beam theorem yield earlier results of this paper. Let the fan beam be infinitely elongated into a strip beam. Then the cross-sectional profile determines the resolution. But this was the result we deduced for strip integration, and it will be seen that the doubly infinite ensemble of measurements made by strip scanning are exactly those made by surveying with a spinning strip beam. For the degeneracy of the strip beam reduces the three independent variables specifying the fan beam position to two. Likewise, when the fan beam is infinitely elongated one way, and infinitely narrow the other, our result shows that spinning the beam will give infinitely sharp resolution. But such a survey gives precisely the measurements obtained by line scanning, which, as known from Section III, uniquely determines the true distribution.

VI. REFERENCES

BOLTON, J. G., and WESTFOLD, K. C. (1951).—*Aust. J. Sci. Res.* A **4**: 476.

BRACEWELL, R. N., and ROBERTS, J. A. (1954).—*Aust. J. Phys.* **7**: 615.

BRACEWELL, R. N. (1955).—*Aust. J. Phys.* **8**: 200.

BRACEWELL, R. N. (1956).—Two-dimensional aerial smoothing in radio astronomy. *Aust. J. Phys.* **9** (3) (in press).

BUSH, V. (1937).—"Operational Circuit Analysis." (Wiley: New York.)

CHRISTIANSEN, W. N., and WARBURTON, J. A. (1953).—*Aust. J. Phys.* **6**: 262.

CHRISTIANSEN, W. N., and WARBURTON, J. A. (1955).—*Aust. J. Phys.* **8**: 474.

COVINGTON, A. E., and BROTEN, N. W. (1954).—*Astrophys. J.* **119**: 569.

DOETSCH, G. (1943).—"Theorie und Anwendung der Laplace-Transformation." p. 298. (Dover: New York.)

HEAVISIDE, O. (1922).—"Electromagnetic Theory." (Benn: London.)

MACHIN, K. C. (1951).—*Nature* **167**: 889.

MILLS, B. Y. (1953).—*Aust. J. Phys.* **6**: 452.

NATIONAL BUREAU OF STANDARDS (1954).—Tables of functions and zeros of functions. Appl. Math. Ser. No. 37.

O'BRIEN, P. A. (1953).—*Mon. Not. R. Astr. Soc.* **113**: 597.

STANIER, H. M. (1950).—*Nature* **165**: 354.

STARR, A. T. (1953).—"Radio and Radar Technique." (Pitman: London.)

WATSON, G. N. (1944).—"Theory of Bessel Functions." 2nd Ed. (Cambridge Univ. Press.)

EMISSION NEBULAE AS RADIO SOURCES

By B. Y. Mills,* A. G. Little,* and K. V. Sheridan*

[*Manuscript received November* 25, 1955.]

Summary

Attempts have been made to detect 14 bright emission nebulae at a wavelength of 3·5 m using a pencil-beam radio telescope with a beamwidth of 50 min of arc. Of these nebulae, six were probably observed in emission, seven were undetectable, and one, NGC 6357, was observed in absorption; radio isophotes were obtained for NGC 2237 and NGC 3372. Radio and optical data have been combined to estimate electron densities, masses, and sometimes the electron temperature of many of the nebulae. Values range from an electron density of 3 cm^{-3} and a mass of 3×10^6 solar masses for the outer regions of the 30 Doradus complex to an average electron density of 500 cm^{-3} and a mass of 20 solar masses for the Orion Nebula. Temperatures generally appear to be in the neighbourhood of 10,000 °K, except in the case of NGC 6357, for which 6500 °K is estimated.

I. Introduction

The role of the interstellar ionized hydrogen in producing metre wave radiation has not yet been satisfactorily determined. For instance, it has been claimed by Scheuer and Ryle (1953) that the gas produces, by virtue of its thermal emission, a band of intense radiation near the galactic centre; but Mills (1955) has shown that, although such a band of emission does exist, it is wider than the figure of 2° given by Scheuer and Ryle and its brightness temperature is too high to be reconciled with observations of the same region at much shorter wavelengths if thermal emission only is assumed. The present investigation was designed to throw some light on the problem by observation of the brightest emission nebulae accessible to the 1500 ft Sydney cross aerial.

Observations of 14 nebulae were made at a wavelength of 3·5 m: six were probably observed in emission, seven were undetectable, and one, NGC 6357, was observed in absorption. It is found that the nebulae emit and absorb radiation in the manner predicted by theory, and do not appear to be associated in any way with the non-thermal galactic radio sources. By combining radio and optical data it has been found possible to estimate electron densities, masses, and sometimes the electron temperature of many of the nebulae. It is not possible from these observations to determine the contribution that the thermal emission of ionized hydrogen makes to the general galactic radiation, but the agreement with theory suggests that detailed surveys at widely different wavelengths will allow this contribution to be deduced, probably in conjunction with the spatial distribution of the hydrogen throughout the Galaxy.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

II. Theory

The emission and absorption of radio waves by the process of free-free transitions in the ionized interstellar gas has been discussed by many authors. We will adopt some results given by Piddington (1951) and derive expressions useful for interpreting the present observations.

Consider a mass of ionized gas with uniform electron temperature T_e and optical thickness τ, situated in front of an extended emitting region of uniform brightness temperature T_b. The apparent excess temperature of the gas mass is given by

$$\begin{aligned}\Delta T &= T_e(1-\mathrm{e}^{-\tau})+T_b\mathrm{e}^{-\tau}-T_b\\ &=(T_e-T_b)(1-\mathrm{e}^{-\tau}). \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)\end{aligned}$$

The gas will therefore appear as an emitter for $T_e>T_b$ and as an absorber for $T_e<T_b$; with equal temperatures it is unobservable. The electron temperature in an H II region is known to be of the order of 10,000 °K, perhaps somewhat higher (see, for instance, Aller 1953). At 3·5 m the apparent brightness temperature near the galactic plane at longitudes within about 15° of the galactic centre is between 8000 °K and 20,000 °K (e.g. Mills 1955); consequently it might be expected that here individual emission nebulae will be observable in absorption, if at all. Elsewhere the brightness temperature is mainly below 2000 °K so that observations of H II regions in emission might be expected. Both these conclusions are supported by the present observations. At shorter wavelengths, with lower values of T_b, the gas might be expected to appear everywhere in emission and, conversely, in absorption at longer wavelengths.

To relate our measurements to the physical properties of the nebulae, we determine the optical depth τ from an expression for the absorption coefficient which is applicable at low electron densities, thus:

$$\left.\begin{aligned}\varkappa &= \frac{n^2}{f^2 T_e^{\,3/2}}\times 9\cdot 70\times 10^{-3}\ln\left(\frac{3kT_e}{2hf}\times 10^{-6}\right),\\ \tau &= \int \varkappa\,\mathrm{d}s,\end{aligned}\right\} \quad \ldots\ldots \quad (2)$$

where n is the electron density, f the frequency in Mc/s, and k and h are Boltzmann's and Planck's constants respectively.

The logarithmic term varies slowly with frequency and temperature; it may be considered invariant over a wide range. Inserting physical constants and assuming an electron temperature of 10,000 °K in the logarithmic term, we have for the optical thickness at wavelengths near 3·5 m,

$$\tau_{(3\cdot 5\mathrm{m})}=\int\varkappa\,\mathrm{d}s=\frac{5\cdot 0\lambda^2}{T_e^{\,3/2}}\times EM, \quad \ldots\ldots\ldots\ldots\ldots \quad (3)$$

where the emission measure, EM, is conventionally defined as $\int n^2\mathrm{d}l$, l being the path length through the gas measured in parsecs, and the wavelength λ is measured in metres. At a wavelength near 10 cm the numerical constant is only slightly different, thus:

$$\tau_{(0\cdot 1\mathrm{m})}=\frac{3\cdot 8\lambda^2}{T_e^{\,3/2}}\times EM. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

The apparent flux density of an emission nebula at constant temperature is obtained from equation (**1**) by applying the Rayleigh-Jeans formula, thus

$$S=\frac{2k}{\lambda^2}(T_e-T_b)\int(1-\mathrm{e}^{-\tau})\mathrm{d}\Omega. \qquad \ldots\ldots\ldots\ldots \quad (\mathbf{5})$$

Two simplified forms arise if τ is either large or small ; thus for large τ,

$$S=\frac{2k}{\lambda^2}(T_e-T_b)\Omega, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{6})$$

and for small τ,

$$S=\frac{2k}{\lambda^2}(T_e-T_b)\Omega\tau. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{7})$$

Substituting equations (**3**) and (**4**) in (**7**) we have two useful expressions for the apparent radio emission of a transparent nebula :

$$\left.\begin{aligned} S_{(3\cdot 5\mathrm{m})}&=3\cdot 3\times 10^{-26}\frac{\overline{EM}}{T^{\frac{1}{2}}}\varphi^2\left(1-\frac{T_b}{T_e}\right),\\ S_{(0\cdot 1\mathrm{m})}&=2\cdot 5\times 10^{-26}\frac{\overline{EM}}{T^{\frac{1}{2}}}\varphi^2\left(1-\frac{T_b}{T_c}\right),\end{aligned}\right\} \qquad \ldots\ldots\ldots \quad (\mathbf{8})$$

where φ is the angular diameter of the nebula measured in degrees.

It is often convenient to assume a model consisting of uniform spherical distribution of fully ionized hydrogen ; we must then integrate the emission measure throughout the sphere. The resulting expressions are, for small τ.

$$\left.\begin{aligned} S_{(3\cdot 5\mathrm{m})}&=2\cdot 2\times 10^{-26}\frac{n^2l}{T^{\frac{1}{2}}}\varphi^2\left(1-\frac{T_b}{T_e}\right),\\ S_{(0\cdot 1\mathrm{m})}&=1\cdot 7\times 10^{-26}\frac{n^2l}{T^{\frac{1}{2}}}\varphi^2\left(1-\frac{T_b}{T_e}\right),\end{aligned}\right\} \qquad \ldots\ldots\ldots \quad (\mathbf{9})$$

where l is the diameter of the sphere, in parsecs.

The total mass in solar mass units of such a model may be estimated from the observations of apparent flux density as follows :

$$M=4\times 10^6 T_e^{\frac{1}{4}}.d^{5/2}.\varphi^{3/2}.S^{1/2}\left(1-\frac{T_b}{T_e}\right)^{-\frac{1}{2}}, \qquad \ldots\ldots \quad (\mathbf{10})$$

where d is the distance to the gas mass in parsecs : the sum total of uncertainties involved does not warrant differentiating between flux densities at different wavelengths. It is interesting that the electron temperature has very little effect on the derived mass.

Equation (**10**) is usually not strictly applicable to a real nebula because electron densities are not uniform. With sufficient data the use of more sophisticated models may be justified, but at present a judicious choice of φ enables the order of mass to be estimated with an accuracy which is reasonably high by astrophysical standards.

Flux density measurements at radio wavelengths are similar in the information they provide to the standard Hα brightness measurements; both are directly proportional to emission measure and are related similarly to electron temperature. However, the radio measurements are unaffected by optical obscuration, which is usually a serious source of uncertainty in the Hα measurements, and, moreover, they yield additional information at metre and decametre wavelengths where most nebulae are more or less opaque. As against this, the resolution of a radio telescope is always vastly inferior to its optical counterpart and usually it is quite incapable of determining the brightness distribution across a nebula. A combination of radio and optical observations is therefore likely to yield considerably more information than either alone; this will become apparent in the subsequent detailed discussion of the observations.

III. Observations

Attempts were made to observe 14 bright emission nebulae including several which have been detected at centimetre wavelengths by Haddock, Mayer, and Sloanaker (1954) and Hagen, McClain, and Hepburn (1954), using the 50 ft N.R.L. paraboloid. Our observations are summarized in Table 1 together with some physical properties of the nebulae obtained by combining these with other radio and optical observations; they are discussed in detail below. Two of the nebulae, NGC 2237 and NGC 3372, were sufficiently large and intense to warrant drawing their radio isophotes. These are shown in Plate 1 in comparison with photographs of the nebulae to the same scale; the isophotes are not corrected for aerial smoothing and the half-power contour of the aerial response pattern is shown. In Figure 1 is reproduced a facsimile of a record taken at the position of NGC 2237, showing the extent of the superimposed random noise fluctuations which limit the accuracy of the radio isophotes.

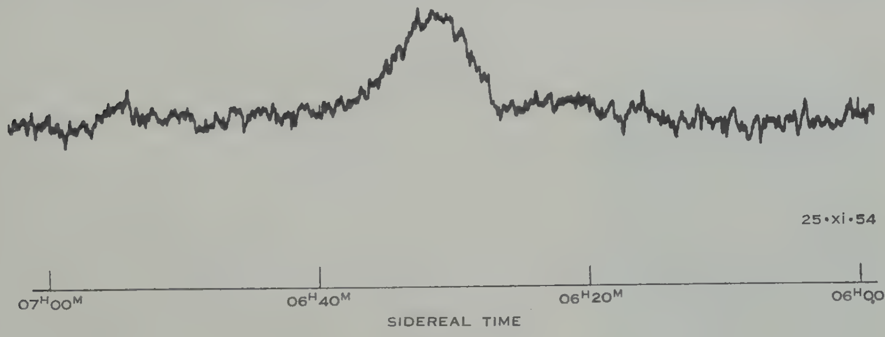


Fig. 1.—Facsimile of a record taken at the position of NGC 2237.

30 Doradus

Details of the observations at 3·5 m of this nebula in the Large Magellanic Cloud have already been published (Mills 1955). The flux density of the whole emission complex is $5 \cdot 5 \times 10^{-25}$ W m^{-2} (c/s)$^{-1}$ and it is estimated that the flux density of 30 Doradus itself is about 3×10^{-25} W m^{-2} (c/s)$^{-1}$, the remainder

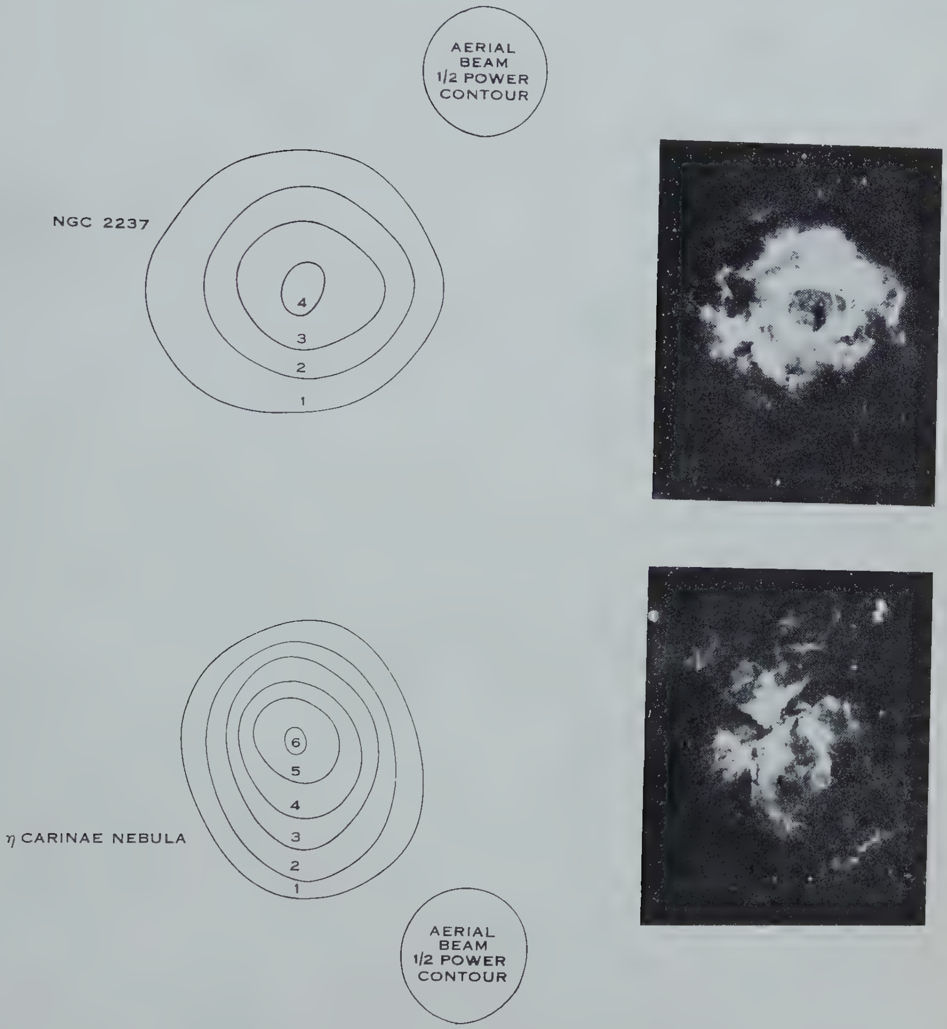
Table 1

EMISSION NEBULAE : FLUX DENSITY AND DERIVED QUANTITIES

Nebula	Apparent Flux Density at 3·5 m (W m^{-2} (c/s)$^{-1}$ $\times 10^{-26}$)	Electron Temperature (°K)	Average Electron Density (cm^{-3})	Mass (⊙)	Remarks
30 Doradus complex	55	10,000*	30 Doradus 40 Neighbouring regions .. 3	4×10^6 3×10^6	Electron density and mass are both very uncertain
NGC 2237	130	10,000	14	8000–10,000	Electron temperature and density are based on optical data : agreement with radio observations is good
NGC 3372 (η Carinae Nebula)	150	10,000*	14	3000	Electron density and mass are probably lower limits
M42 (Orion Nebula)	37	10,000*	500	20	Electron density and mass are derived from flux density measurements at different wavelengths.
IC 2177	15	—	—	—	Identification uncertain
NGC 2264	8	—	—	—	Identification uncertain
NGC 3324	<20	10,000*	<40	<7000	
NGC 6188	<10	~10,000	—	—	Undetectable although all are strong sources at centimetre wavelengths ; electron temperatures must be close to the background temperature, in all cases of the order of 10,000 °K
NGC 6334	<20	~10,000	—	—	
M16	<20	~10,000	—	—	
M17	<10	~10,000	—	—	
M8	<30	~10,000	—	—	
M20	<50	~10,000	—	—	
NGC 6357	−85	6,500	37	1500	Observed in absorption

* Value assumed.

EMISSION NEBULAE AS RADIO SOURCES



Comparisons of the observed 3·5 m radio isophotes and photographs of the nebulae NGC 2237 (Mount Wilson and Palomar Observatories photograph) and NGC 3372 (Harvard College Observatory, Boyden Station photograph).

being made up by the neighbouring emission regions mainly to the south. The flux density of 30 Doradus may be estimated independently on the assumption that at a wavelength of 3·5 m the nebula is opaque, has an angular size of 15 min of arc, an electron temperature of 10,000 °K, and the background temperature is 500 °K. The resulting flux density is $3 \cdot 2 \times 10^{-25}$ W m^{-2} (c/s)$^{-1}$ in good agreement with the radio data, suggesting that these assumptions are probably substantially correct. Measurements of the nebula at a wavelength of 50 cm have been made by Piddington and Trent (1956); they find a source of flux density $4 \cdot 5 \times 10^{-24}$ W m^{-2} (c/s)$^{-1}$ at the position of the emission complex. It should eventually be possible to combine the radio data to derive reliable values for electron density and temperature in the nebula and surrounding regions. This would be quite a straightforward procedure if the Hα brightness distribution were available, but in the absence of such data there is little point in a detailed analysis. Both radio observations could be reproduced by a model consisting of a dense spherical concentration 200 pc in diameter and with uniform electron density of 40 cm^{-3}, contiguous with a uniform spheroidal concentration 600 pc in diameter and 200 pc thick with an electron density of 3 cm^{-3}, all at a uniform electron temperature of 10,000 °K. The corresponding masses of the gas concentrations are 4×10^6 and 3×10^6 solar masses.

NGC 2237

This nebula is easily resolved and isophotes are shown in Plate 1; maximum emission occurs at a position of $06^{\mathrm{h}}\,29^{\mathrm{m}} \cdot 6$, $+5°\,01'$ (1950). Minkowski (1949) has studied the nebula and deduced the physical state of the gas: he has recently revised this result (Minkowski 1955) and, based on his new data, a model has been assumed which consists of a uniform sphere 35 pc in diameter with an electron density of 14 cm^{-3} at a temperature of 10,000 °K located at a distance of 1400 pc. At decimetre wavelengths this model would give a flux density of $2 \cdot 2 \times 10^{-24}$ W m^{-2} (c/s)$^{-1}$ and at 3·5 m an apparent flux density of $1 \cdot 6 \times 10^{-24}$ W m^{-2} (c/s)$^{-1}$, in reasonably good agreement with the present observations and also with a measurement of Ko and Kraus (1955), who found a flux density of $2 \cdot 0 \times 10^{-24}$ W m^{-2} (c/s)$^{-1}$ at a wavelength of about 1 m. However, Piddington and Trent (1956) quote a flux density of 4×10^{-24} W m^{-2} (c/s)$^{-1}$ at a wavelength of 50 cm for a source which they identify with the nebula. If this is correct the model assumed must be wrong. The occurrence of small dense regions within the nebula could perhaps explain the discrepancy. The mass of the model is 8000 solar masses and the mass corresponding to the measurement of Piddington and Trent is about 10,000 solar masses. The shape of the isophotes in Plate 1 is consistent with an absence of emission from the centre of the nebula, but the resolution is inadequate to establish this beyond doubt.

NGC 3372 (η Carinae Nebula)

Of the emission nebulae observed, this is the most prominent radio source and it is also easily resolved by the aerial. Isophotes are shown in Plate 1: the position of maximum emission is $10^{\mathrm{h}}\,43^{\mathrm{m}} \cdot 4$, $-59°\,31'$ (1950). Unfortunately no optical data relating to emission measures or brightness distributions are available and it could not be separately resolved in the 50 cm survey by Piddington and Trent; consequently no reliable estimates can be made of the physical state of the system. The central concentration is clearly much more pronounced than in NGC 2237, which is evident also from the photographic appearance. If the nebula is assumed transparent at 3·5 m with electron temperature 10,000 °K, background temperature 2000 °K, and at a distance of 1100 pc equal to that of the presumed exciting star, η Carinae, we find an r.m.s. electron density of 14 cm^{-3} and a corresponding mass of about 3000 solar masses. These are likely to be lower limits, however, because it is probable that the central portion of the nebula is opaque at this wavelength.

M42 (Orion Nebula)

This nebula offers perhaps the best opportunities for a detailed analysis since it has been studied extensively both at radio and optical wavelengths. However, optical measurements appear to differ widely, presumably because of the great variations in density across the nebula and the large amount of obscuration in the region. It is difficult to assess the significance of these differences, consequently we will in this paper attempt to deduce some properties of the nebula almost entirely from the radio observations.

The observations of the nebula are rather more uncertain than those already described because of the possibility of interference from the intense radio source Taurus-A which culminates at nearly the same time. However, it is considered that such interference is not likely to be serious. The position obtained is $05^{\mathrm{h}}\,32^{\mathrm{m}}\cdot6$, $-5°\,23'$, displaced by a few minutes of arc from the position given by Haddock, Mayer, and Sloanaker (1954) observing at 9·4 cm; but this is to be expected as the electron density is not constant over the nebula. If it is assumed that the nebula is optically thick at 3·5 m and has an electron temperature of 10,000 °K, an "effective angular diameter" may be estimated from the flux density measurement, by substitution in equation (**6**) of Section II. This diameter is 16 min of arc, which is consistent with a slight widening of the apparent response pattern of the aerial from 50 to about 53 min of arc when observing the source. Accordingly we shall assume a uniform spherical model of diameter 1·4 pc, corresponding to the nebula distance of 300 pc. The flux-density at 9·4 cm is $4\cdot5\times10^{-24}$ W m^{-2} $(\mathrm{c/s})^{-1}$ leading to an electron density of about 500 cm^{-3} and a mass of about 20 solar masses.

IC 2177

The radio emission shows considerable structure near this emission complex and until isophotes of the whole region become available it is not possible to make any certain identifications. However, there appears to be a source centred on $07^{\mathrm{h}}\,03^{\mathrm{m}}\cdot6$, $-10°\,40'$ (1950) extending for about a degree in a roughly north-south direction; it may probably be identified with the main nebulosity of the area. An estimate of the physical state of the system is not possible with the present inadequate data.

NGC 2264

This is a relatively faint nebula surrounding a cluster of stars. The radio source is very weak and is in a region of irregular background variation so that the identification is uncertain. Furthermore, insufficient observations are available at adjacent declinations to fix the position with any accuracy: it is estimated as $06^{h}\,36^{m}\cdot5 \pm 1^{m}$, $+10^{\circ} \pm \frac{1}{2}^{\circ}$ (1950), which is sufficiently close to the centre of the nebula to warrant suggesting a possible identification. The estimated flux density is also uncertain because of difficulties in allowing for the angular size of a weak source. A diameter of $\frac{1}{2}^{\circ}$ was assumed, comparable with the size of the nebula.

NGC 3324

This is a small and not particularly bright nebula in a region of fairly high and irregular background temperature; failure to detect it is therefore not surprising. Assuming an angular size of 15′, a distance of 3000 parsecs corresponding to that of the exciting star GC 14621, an electron temperature of 10,000 °K, and a background temperature of 3000 °K, we may deduce from equation (9) that the electron density is less than 40 cm^{-3} and the mass less than 7000 solar masses.

NGC 6188, NGC 6334, M16, M17

These nebulae are all in regions with background temperatures of the order of 10,000 °K. In no case could excess radiation be detected and in one case, M17, there is some suggestion of a slight absorption, but further observations are required to verify this. All the nebulae have been observed at shorter wavelengths by one or more of several observers (Haddock, Mayer, and Sloanaker 1954; Hagen, McClain, and Hepburn 1954; McGee, Slee, and Stanley 1955): it is possible to infer from the failure to observe them at 3·5 m that electron temperatures are generally in the neighbourhood of 10,000 °K and that there is little, if any, non-thermal radiation emitted by the nebulae.

M8, M20

Both these nebulae, particularly M20, occur close to an intense non-thermal source, 17–2A (Mills 1952), which has a position of $17^{h}\,57^{m}\cdot4$, $-23^{\circ}\,25'$ (1950). No evidence for any effect, either in absorption or emission, can be found at the nebulae positions. An association with the source 17–2A therefore appears unlikely.

NGC 6357

This is the only nebula studied which shows clear evidence of an absorption effect. Tracings of records made on adjacent declinations straddling the nebula are shown in Figure 2: the position of maximum absorption is estimated as $17^{h}\,22^{m}\cdot0$, $-34^{\circ}\,09'$ (1950) in very good agreement with the position of the nebula given by Bok, Bester, and Wade (1954), that is $17^{h}\,21^{m}\cdot7$, $-34^{\circ}\,07'$ (1950). Measurements of the nebula at a wavelength of 9·4 cm have been made by Haddock, Mayer, and Sloanaker who quote the identical position that we obtain. If we assume an angular size of $63' \times 43'$ as given by Bok, Bester, and Wade and a flux density at 9·4 cm of $4\cdot0 \times 10^{-24}\ \mathrm{W\,m^{-2}\,(c/s)^{-1}}$, it follows that at a wavelength of 3·5 m the nebula must be optically thick, that is, $e^{-\tau}\simeq 0$. In these circumstances we may use equation (**6**) to derive the electron temperature, but it is more illuminating to analyse the data directly in terms of brightness temperatures.

It is clear that the electron temperature of a nebula observed in absorption must normally be less than the brightness temperature measured at its position ;* in the present case the brightness temperature is 7800 °K. The estimated maximum calibration error in this part of the sky is ±20 per cent., or 1600 °K ; accordingly we conclude that the electron temperature of the nebula is probably less than 7800 °K and almost certainly less than 9400 °K. An estimate of the actual temperature may be obtained from the following equation, derived from equations (**1**) and (**5**) of the preceding section :

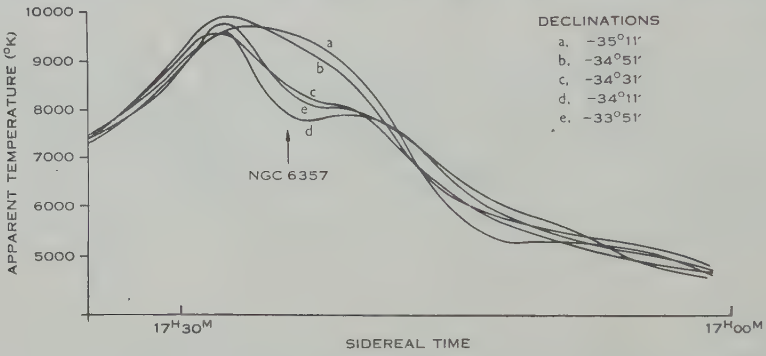


Fig. 2.—Tracings of records at declinations straddling the nebula NGC 6357, superimposed to display the effect of absorption.

$$T_e=T_a-(T_0-T_a)\left(\frac{\Omega_0}{\Omega_n(1-e^{-\tau})}-1\right)-\alpha D, \quad \ldots\ldots\ldots (\mathbf{11})$$

where T_a is the brightness temperature at the nebula position ; T_0 is the temperature which would have been observed in the absence of the nebula, estimated by interpolation ; Ω_0 is the equivalent solid angle of the nebula when smoothed by the aerial beam ; Ω_n is the actual solid angle of the nebula ; α is a measure of the density of galactic radio emission in units of temperature per unit distance measured in the direction of the nebula ; and D is the distance to the nebula.

We take the following values :

$T_a=7800$ °K,
$T_0=8800$ °K,
$\Omega_0=3\cdot6\times10^{-4}$ steradians,
$\Omega_n=1\cdot9\times10^{-4}$ steradians (corresponding to a size of $63'\times43'$),
$e^{-\tau}\simeq0$ (consistent with the measurements of flux density at 9·4 cm),
$\alpha\simeq0\cdot5$ °K/parsec (estimated from the integrated brightness through the Galaxy),
$D=1000$ parsecs (distance of the assumed exciting star −34°11671).

---

* If both the nebula and background are irregular with comparable scale factors this conclusion is not necessarily true, but it appears safe to ignore such a possibility here.

The corresponding electron temperature is 6500 °K, substantially lower than the value of 10,000 °K usually assumed for such nebulae. It would be interesting to attempt to check this abnormally low value by optical spectral measurements. With this value of T_e and the flux density obtained at 9·4 cm, we find an electron density of 37 cm^{-3} and a mass of 1500 solar masses. With these parameters it is surprising that the nebula is not more prominent optically ; probably the large amount of obscuration in the region is responsible.

IV. Conclusions

A significant deduction from the observations is that ordinary emission nebulae emit radio waves as the result of their high electron temperature by the process of free-free transitions. There is no indication whatsoever that they have an association with the more intense non-thermal galactic radio sources. Accordingly it appears quite safe to apply the well-known theory of this process in determining the properties of the ionized interstellar gas. Detailed high resolution surveys of the Milky Way at widely different frequencies should enable the distribution and the physical state of the gas to be determined. A crude distance scale is available for individual distant gas concentrations by estimation of the background temperature, T_b, from observations at different wavelengths ; this may well be of use in determining a three-dimensional model of the distribution. It is essential that all surveys should be complete and of high resolution, as a casual inspection of our records reveals a complicated structure along the length of the southern Milky Way ; until the records are analysed in detail it is not possible to estimate the resolution actually required.

Finally, the importance of quantitative optical studies of emission nebulae is obvious ; particularly in the extremely rich southern fields. Apart from the possibility of deriving physical properties of the nebulae, independent estimates of optical extinction may also arise from comparisons between the radio and optical data.

V. References

Aller, L. H. (1953).—" Astrophysics : The Atmospheres of the Sun and Stars." (Ronald Press : New York.)

Bok, B. J., Bester, M. J., and Wade, C. M. (1954).—Catalogue of HII regions in the Milky Way for longitudes 250°–355°. Harvard Rep.

Haddock, F. T., Mayer, C. H., and Sloanaker, R. M. (1954).—*Nature* **174** : 176.

Hagen, J. P., McClain, E. F., and Hepburn, N. (1954).—*Astr. J.* **59** : 323.

Ko, H. C., and Kraus, J. D. (1955).—*Nature* **176** : 221.

McGee, R. X., Slee, O. B., and Stanley, G. J. (1955).—*Aust. J. Phys.* **8** : 347.

Mills, B. Y. (1952).—*Aust. J. Sci. Res.* A **5** : 266.

Mills, B. Y. (1955).—*Aust. J. Phys.* **8** : 368.

Minkowski, R. (1949).—*Publ. Astr. Soc. Pacif.* **61** : 151.

Minkowski, R. (1955).—Gas dynamics of cosmic clouds. I.A.U. Symposium Ser. No. 2, p. 3. (North Holland Publishing Co. : Amsterdam.)

Piddington, J. H. (1951).—*Mon. Not. R. Astr. Soc.* **111** : 45.

Piddington, J. H., and Trent, G. H. (1956).—*Aust. J. Phys.* **9** : 74.

Scheuer, P. A. G., and Ryle, M. (1953).—*Mon. Not. R. Astr. Soc.* **113** : 3.

SPORADIC E AT BRISBANE

By J. A. Thomas*

[*Manuscript received December* 29, 1955]

Summary

Further data are presented in support of McNicol and Gipps's classification of two types of sporadic E at Brisbane (lat. 27·5 °S.), namely, sequential type (E_{ss}) and constant-height type (E_{sc}).

Information is given on the occurrence, critical and blanketing frequencies, reflection coefficients, range spreading, vertical movements, and lateral extent of E_s patches.

No correlation is found with sunspot cycle, meteor occurrence, or thunderstorm activity. Scattering from a turbulent medium is not responsible for E_{sc} at Brisbane. Tidal movement is regarded as the most likely cause of both types of E_s.

A method of obtaining information about D region absorption is briefly discussed.

I. INTRODUCTION

The sporadic E (E_s) region has been studied in some detail at Brisbane and a previous paper by McNicol and Gipps (1951) has given the broad outlines of its characteristics. This work has been continued and a more detailed examination of the $P'f$ records, now available for a period of 11 years, has led to some modification and enlargement of the conclusions previously drawn.

McNicol and Gipps have shown that a detailed examination of the $P'f$ records often allows a classification of the E_s layers represented on the records into two distinct types, which vary in respect to the phenomena associated with their appearance, and with respect to their blanketing properties. These two types were given the names " sequential " E_s (E_{ss} or E_{2s}) and " constant-height " type E_s (E_{sc}). Typical examples of these two types of E_s are shown in Plate 1 and Figure 1. Sequential E_s is characterized by its regular decrease in range and increase in critical and blanketing frequencies. Figure 2 gives the average variation of critical and blanketing frequencies with group height for this type of layer for December and January of the years 1950–53. An important secondary characteristic of E_{ss} is the fact that its blanketing frequency is always high—by the time the layer reaches the 120 km level the blanketing frequency is nearly always greater than 4 Mc/s. It is only during the last stages in the life of an E_{ss} region which has persisted for some hours that the strong blanketing by the region may cease.

The constant-height type of E_s layer gives a $P'f$ trace which is often patchy in nature and which appears in its final position, usually without any changes of height. It causes very weak blanketing which very rarely extends beyond 3·5 Mc/s. E_{sc} may occur simultaneously with E_{ss}, and may be above or (more frequently) below this region.

---

* Physics Department, University of Queensland, Brisbane.

It is found that above about 120 km one can determine with certainty, by examination of a succession of records, which type of E_s is present; the majority of such records are, in fact, due to sequential E_s in its early stages. At lower levels the classification becomes increasingly difficult and often impossible. Most of the E_s recorded at Brisbane is at or below the 120 km level, and the majority of the work described below refers to E_s occurring at or below the 120 km level. Specific mention is made when this is not the case.

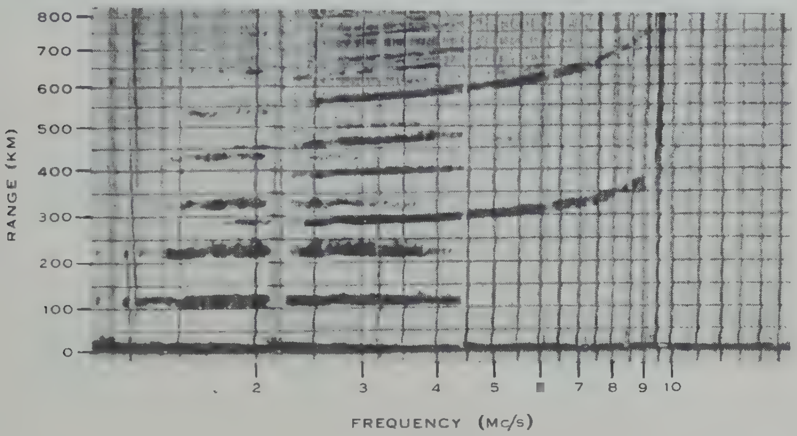


Fig. 1.—A typical $P'f$ record of E_{ss}.

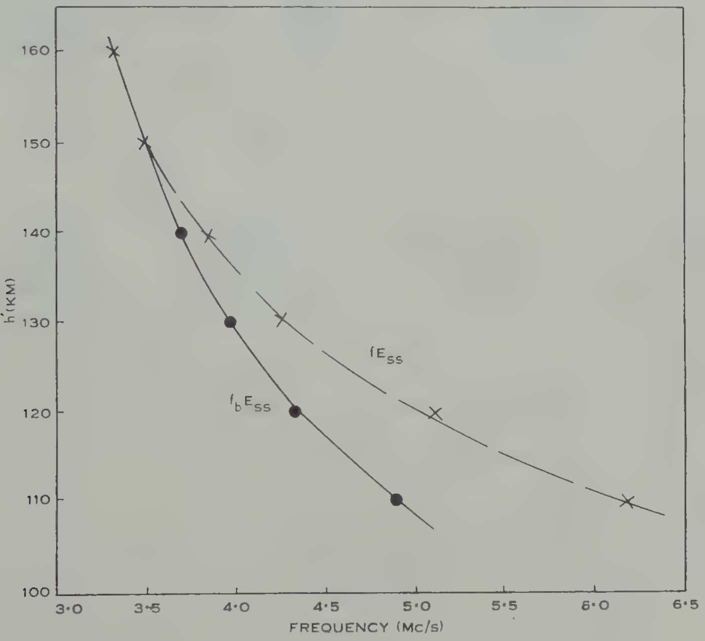


Fig. 2.—Variation of fE_s and f_bE_s with height for sequential E_s.

II. Equipment Used

The automatic $P'f$ equipment in use and the scaling procedures adopted have been fully described by McNicol and Gipps (1951). In the period June 1943 to July 1947, the equipment operated over the range $2\cdot2$–13 Mc/s; from July 1947 onwards, the range was increased to cover 1–16 Mc/s. The earlier

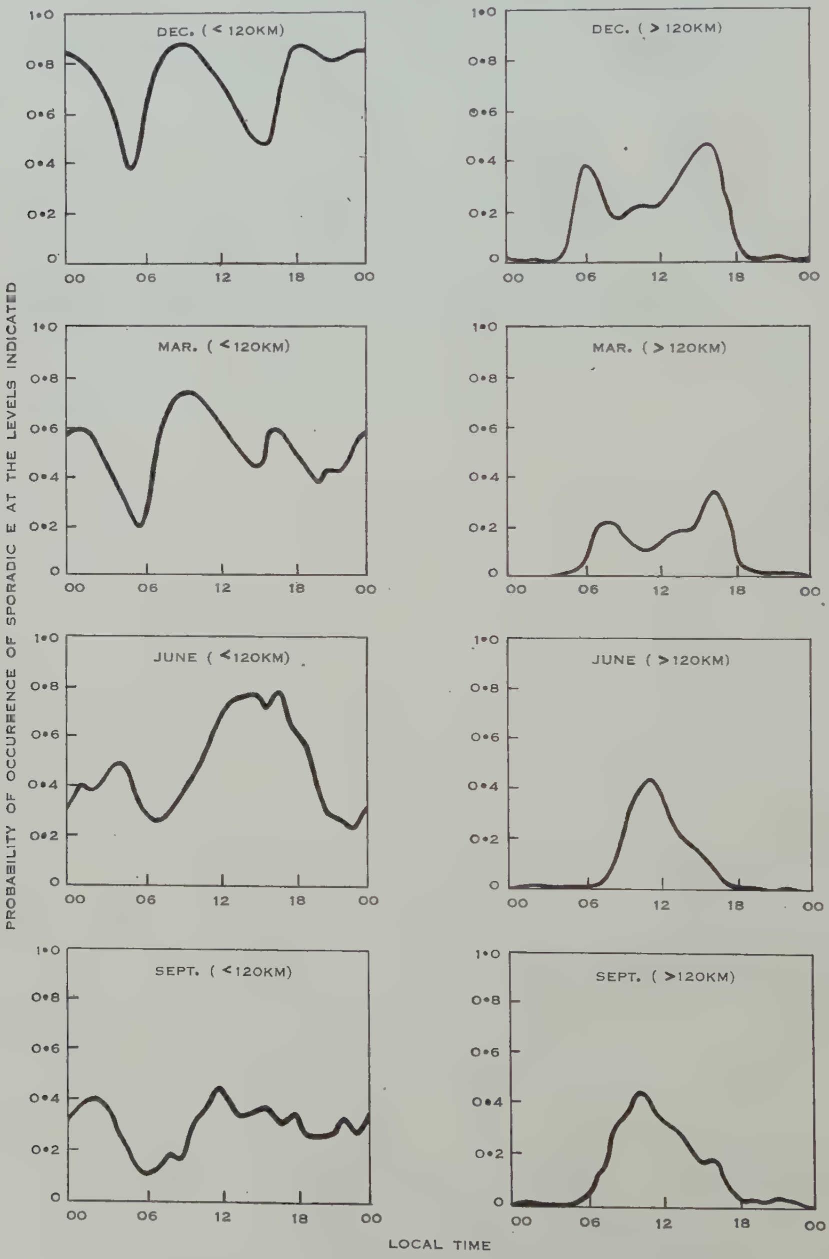


Fig. 3.—Seasonal-diurnal probability of occurrence of E_s echoes.

equipment radiated about 200 W in a 100 μsec pulse and the later equipment about 500 W in a 70 μsec pulse. Due to variations in radiated power, receiver sensitivity, radio interference, and tuning, the overall sensitivity of the recorder

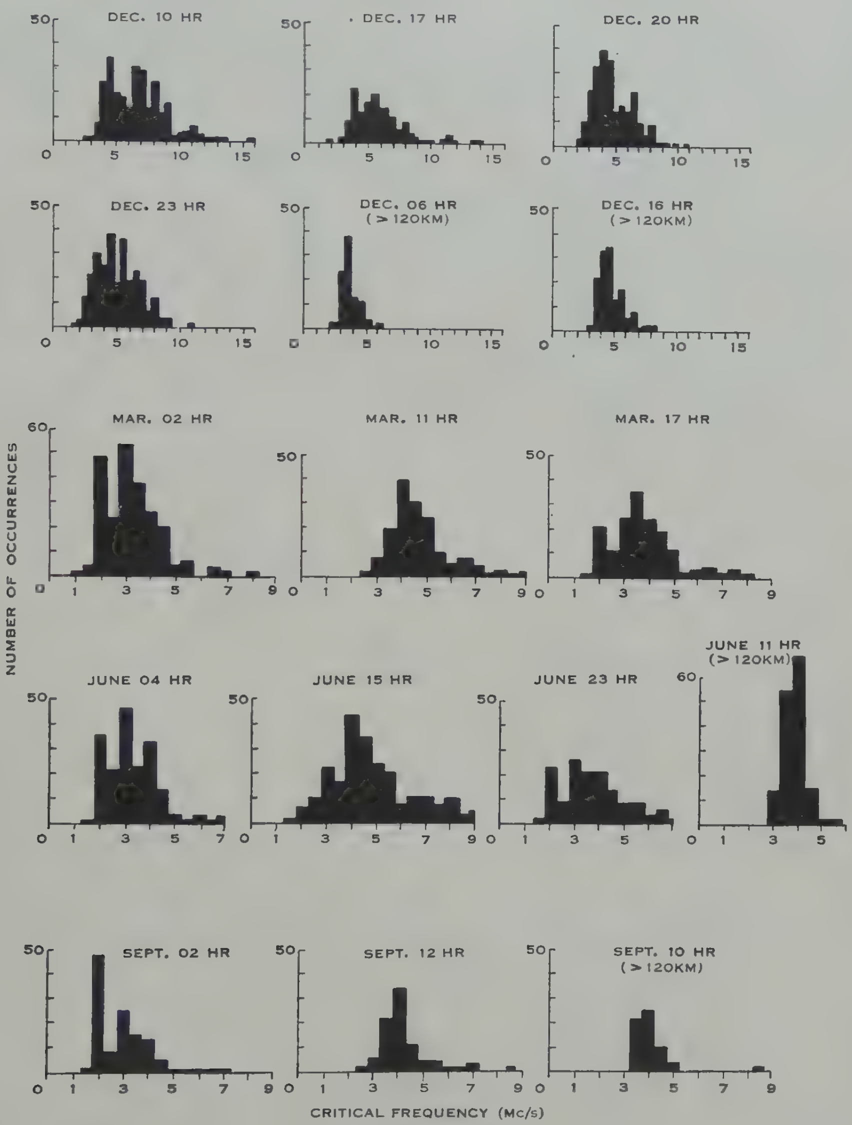


Fig. 4.—Frequency distributions for fE_s.

is by no means constant throughout the band of frequencies covered. A rough estimate shows that an effective reflection coefficient of about 0·003 is necessary before an echo will be recorded by the later equipment; a slightly higher value was necessary for the older equipment.

Some use has also been made of recordings of echoes occurring at fixed frequencies of 2·28, 3·84, and 5·80 Mc/s (*P′t* recordings); these equipments operate at a higher transmitter power output of 1 kW and will record echoes from a region with effective reflection coefficient down to 0·001. The confinement to a single frequency makes possible a number of subsidiary experiments such as the measurement of the direction of arrival of the downcoming waves (Thomas and McNicol 1955), phase-path measurements, and the measurement of the relative intensity of echoes by the swept-gain technique* (McNicol, Webster, and Bowman 1956).

III. Occurrence of E_s and E_2

The probability of occurrence of E_s above and below the 120 km level is shown for four representative months of the year in Figure 3. These figures are derived from *P′f* records and are therefore subject to the limited frequency coverage of the equipment.

Histograms are shown in Figure 4 of the distribution of critical frequencies of the E_s region for various seasons and hours. The fact that the distributions of Figure 4 fall off markedly at a frequency above the lowest frequency recorded (1 Mc/s) suggests that a distinction should be made between the type of E_s recorded between 1 and 16 Mc/s at Brisbane and the type recorded at much lower frequencies elsewhere in the world (Brown and Watts 1950; Watts and Brown 1951; Helliwell 1952; Lindquist 1953). If this is valid, the occurrence figures are not seriously affected by the limited frequency coverage of the recorder.

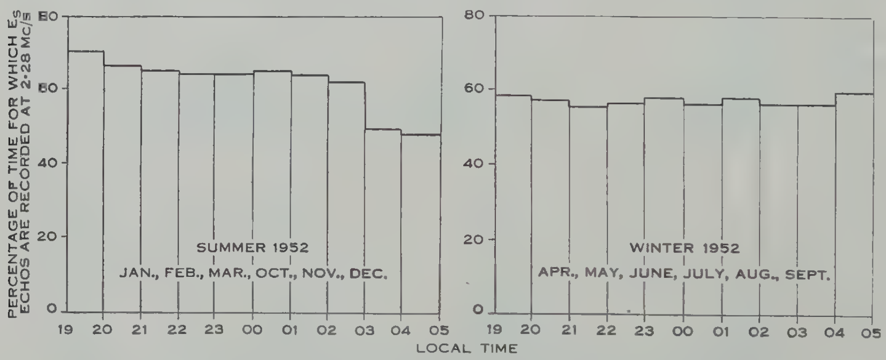


Fig. 5.—Summer and winter night-time E_s occurrence at 2·28 Mc/s.

Night-time *P′t* records taken at 2·28 Mc/s with somewhat higher transmitter power and receiver sensitivity (about 10 dB overall improvement on *P′f* equipment) show E_s echoes for about 60 per cent. of the night. The diurnal distribution of occurrence is fairly uniform throughout most of the night but decreases in the last few hours before dawn in the summer months (Fig. 5). The overall occurrence in summer is greater than that in winter.

---

* The receiver gain is uniformly reduced by 80 dB in a period of 2 min, the process being repetitive. The stronger the echo, the longer it lasts in any 2 min period.

Since sequential type E_s originates in stratification between the E_1 and F_1 regions, the frequency of occurrence of such E_2 echoes is of some importance in any discussion of the nature of E_{ss}. The seasonal-diurnal occurrence of E_2 echoes is shown in Figure 6.

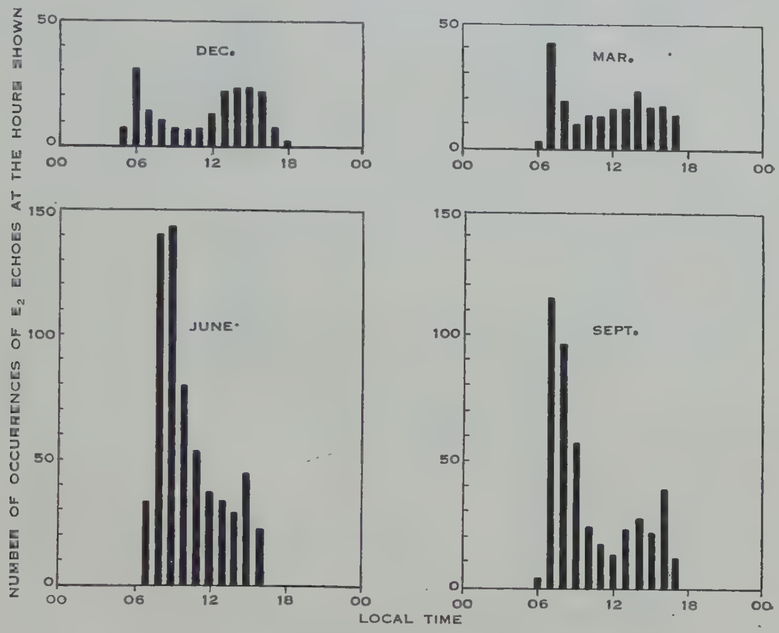


Fig. 6.—Seasonal-diurnal occurrence of E_2 echoes.

IV. Critical and Blanketing Frequencies

While it is generally possible to calculate the mean value of fE_s for any particular hour or season, this is often impossible for f_bE_s, due to the facts that (*a*) the blanketing frequency may be less than the lower limit of recording, and (*b*) the value of f_bE_s may be doubtful owing to overlap of normal E and E_s echoes, or because f_xF_2 may be less than the blanketing frequency (giving rise to complete blanketing). To overcome this difficulty, use is made of the fact that for a normal distribution of values, the mean value of f_bE_s or fE_s increases with the number of occasions on which f_bE_s or fE_s exceeds an arbitrary frequency near the median value. That the distribution of f_bE_s and fE_s is, in fact, an approximately normal one is shown in Figures 7 and 4 respectively. The arbitrary comparison frequencies chosen are 4 and 5 Mc/s respectively; these values are the most suitable for seasonal comparison. For convenience in comparison of the behaviour of critical and blanketing frequencies, the same procedure is used for both quantities.

McNicol and Gipps (1951) plotted histograms on this basis for the various months of the year and found a summer maximum of both critical and blanketing frequencies and a small winter sub-maximum of critical frequency alone. Rather more information may be obtained from seasonal-diurnal plots, and these are

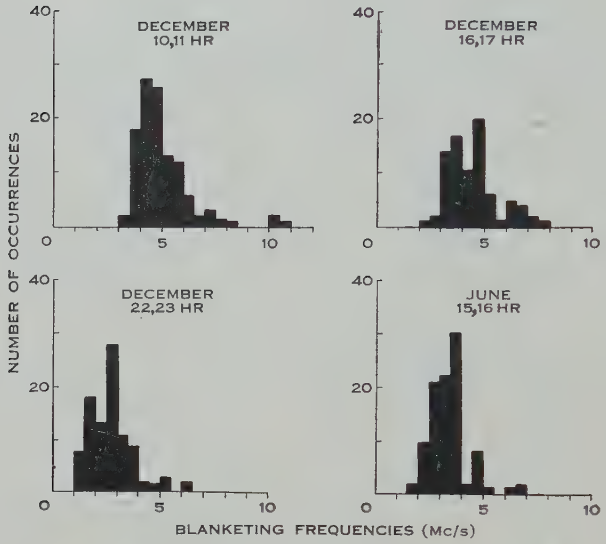


Fig. 7.—Frequency distributions for f_bE_s.

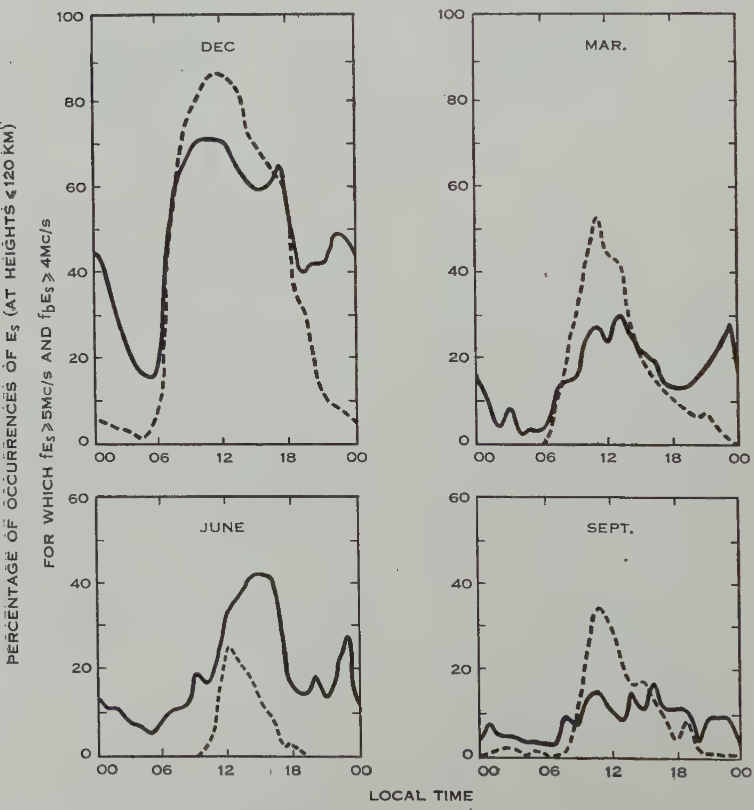


Fig. 8.—Seasonal-diurnal percentage occurrence of $fE_s \geqslant 5$ Mc/s and $f_bE_s \geqslant 4$ Mc/s.

shown for four representative months (averaged over 10 years) in Figure 8. Examination of Figure 8 brings out several points :

(*a*) In all seasons there is a dawn minimum of critical and blanketing frequencies ;

(*b*) In December there are three peaks in critical frequency, the first two at about 1000 hr and 1700 hr being accompanied by peaks in blanketing frequency, but the last at 2230 hr occurring at a time of low blanketing frequency ;

(*c*) In March there are only two predominant peaks in critical frequency at about 1200 hr and 2300 hr of which only the first is accompanied by strong blanketing ;

(*d*) In June there are two peaks of critical frequency at about 1500 hr, and 2300 hr while the blanketing frequency reaches a maximum at 1200 hr ;

(*e*) In September the critical frequency does not vary much through the day but the blanketing frequency reaches a maximum at 1000 hr.

The range of frequencies for which echoes are received from both the E_s and higher regions is given by $fE_s - f_bE_s$. Some indication of the increase in this quantity as sequential E_s moves downwards has been given in Figure 2.

TABLE 1

PARTIAL TRANSMISSION RANGES

Time	Predominant E_s	$fE_s - f_bE_s$ (Mc/s)
December 0700 hr (>120 km) . .	E_{ss}	0·25
December 1100 hr	E_{ss}	1·0
December 2200 hr	E_{sc}	2·7
March 1000 hr	E_{ss}	0·8
March 2300 hr	E_{sc}	1·9
June 1500 hr	E_{sc}	1·6

Average values for various times are given in Table 1. Sequential E_s has an obviously smaller range of partial transmission and reflection than E_{sc}. Rawer (1949), Briggs (1951*b*), and Chatterjee (1953) have made more detailed measurements of this range of partial transmission and reflection. Measurements made at Brisbane by introducing known changes of sensitivity in the *P′f* recorder (Thomas, Svenson, and Brown 1956) have shown that, within the limits of normal scaling measurements (0·1 Mc/s), $fE_s - f_bE_s$ is independent of recorder sensitivity over a range of 30 dB ; such fluctuations as do occur with change in sensitivity are no greater than those commonly occurring on fixed sensitivity.

The recorder sensitivity measurements also showed that at Brisbane, if E_s is of sufficient strength to record at all, its critical frequency is independent of equipment sensitivity for *both* E_{ss} and E_{sc}.

V. Multiple-hop Critical Frequency

In summer it is sometimes observed that the critical frequency of the trace representing two hops between ground and E_s is somewhat higher than that of the trace representing a single hop between ground and E_s (Plate 2, Fig. 1). If we denote these frequencies by f_mE_s and fE_s, then Figure 9 shows the diurnal variation of the occurrences of $f_mE_s > fE_s$ in December. There is a negligible number of occurrences of this phenomenon in winter.

Dieminger (1951) has discussed such records and states that the slight height rise associated with the higher critical frequency of the second echo is consistent with the hypothesis of ground scatter and oblique incidence propagation via the E_s region. Detailed measurements of records such as that shown in Plate 2, Figure 1, indicate, however, that the increases in range are much smaller than is necessary to fit Dieminger's theory. It is certainly true that the second echo generally shows evidence of range spreading, but it is believed that this is more the result of the ionosphere behaving as a diffuse reflector giving a totally different type of illumination for the second hop, rather than ground scatter. Dieminger found it necessary to use very high powers to consistently observe ground scatter; at Brisbane, $2F_2$ scattered echoes such as he describes are never observed, and it is unlikely that such scattered echoes should be observed on day-time $2E_s$ echoes, and yet not on night-time $2F_2$ echoes.

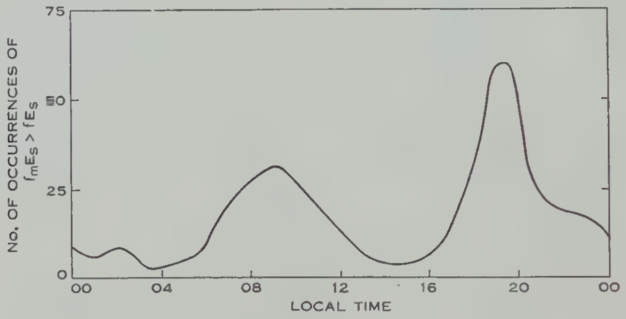


Fig. 9.—December variation of occurrence of $f_mE_s > fE_s$.

Whatever explanation is finally advanced, it would seem that sequential E_s in its later stages (i.e. having dropped to its basic level and commenced to decay) is particularly favourable for the occurrence of such a phenomenon. It is believed (see Section XI (*a*)) that the sequential region at this stage consists of thin disk-like clouds which cannot be observed at any high degree of obliquity.

Records are often obtained with traces corresponding to a ray path from ground to F region to top of E_s region to F region and back to ground. These are the so-called M echoes. A typical summer and winter distribution of the occurrence of M echoes is given in Figure 10 for December and June, 1945. A comparison of these diagrams with Figure 3 shows an obvious correlation between the presence of E_s and the occurrence of M echoes, namely, M echoes are liable to be observed whenever E_s is observed below 120 km.

Sometimes records are made in which an *M* echo is observed reflecting at higher frequencies than the E_s layer itself ; indeed, the E_s trace may be entirely absent from the record. Thus the " top " of an E_s region may, on occasion, be a much better reflector of radio waves than the lower boundary of the region.

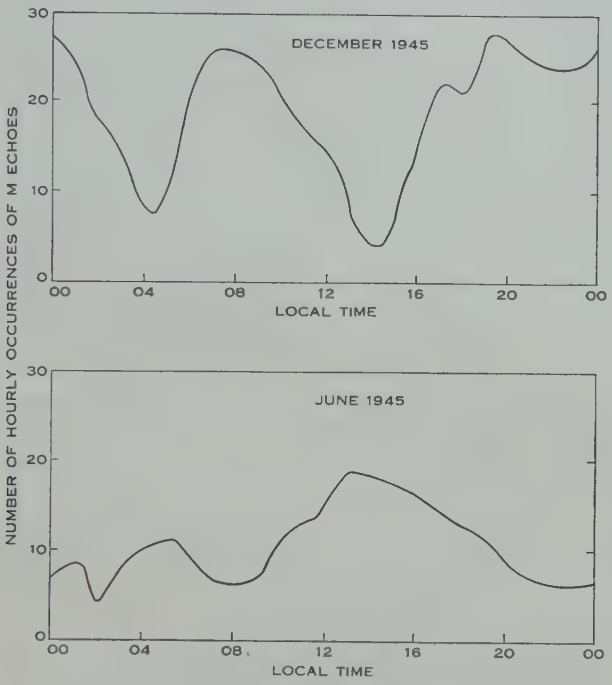


Fig. 10.—Summer and winter *M*-echo occurrence.

VI. Virtual Height of E_s Echoes

Using the published values of $h'E_s$ taken from $P'f$ records over 10 years, the diurnal variation of $h'E_s$ has been found and is shown in Figure 11. There is a large semi-diurnal variation in summer, autumn, and winter, but the spring equinox shows only a very ragged small diurnal variation. The phase of the height oscillation in June leads that in December by about 7 hr. The mean level of the E_s reflection shows a considerable seasonal fluctuation, being 5 km lower in winter than in summer.

A frequency plot of the occurrence of night-time E_s echoes at various heights (measured at 2·28 Mc/s to the nearest km by the method outlined by Thomas and Svenson (1955)) is shown in Figure 12 for the combined months of December 1952 and January 1953. The density of shading increases with increasing frequency of occurrences of recorded height values. The intermingling of the two types of E_s is quite obvious : a constant-height type layer at about 110 km is superimposed on the last stages of an afternoon sequential type layer, becoming prominent at about 2100 hr.

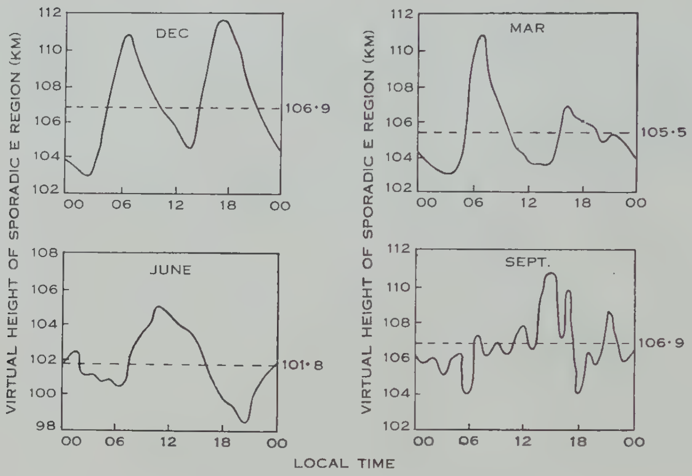


Fig. 11.—Seasonal-diurnal variations of $h'E_s$.

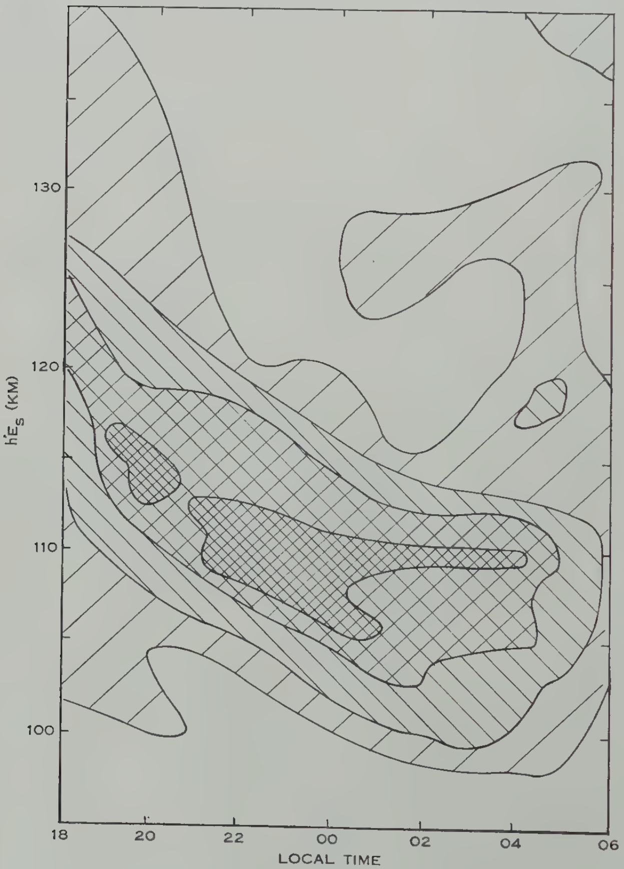
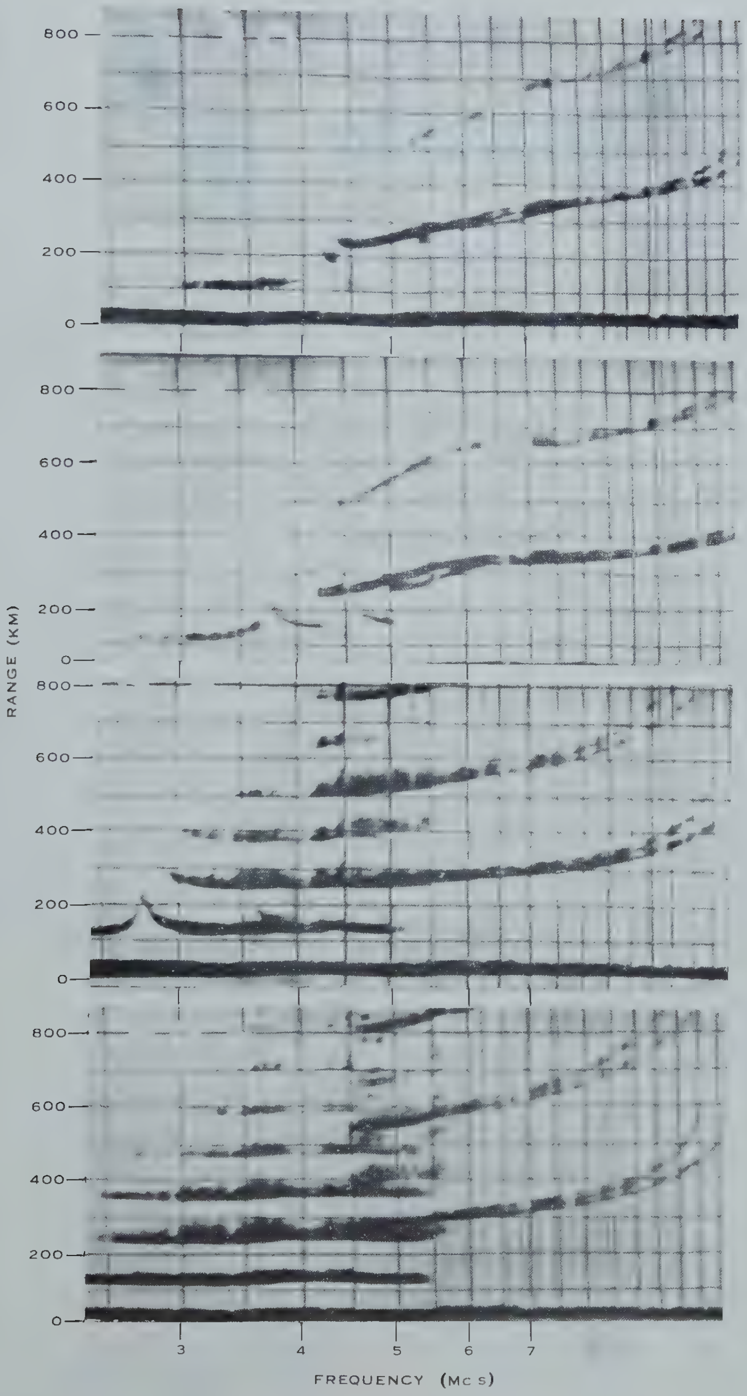


Fig. 12.—Mass plot of $h'E_s$ (measured at 2·28 Mc/s) for summer 1952–53.

SPORADIC E AT BRISBANE



A series of $P'f$ records showing the formation sequence of E_{ss}.

SPORADIC E AT BRISBANE

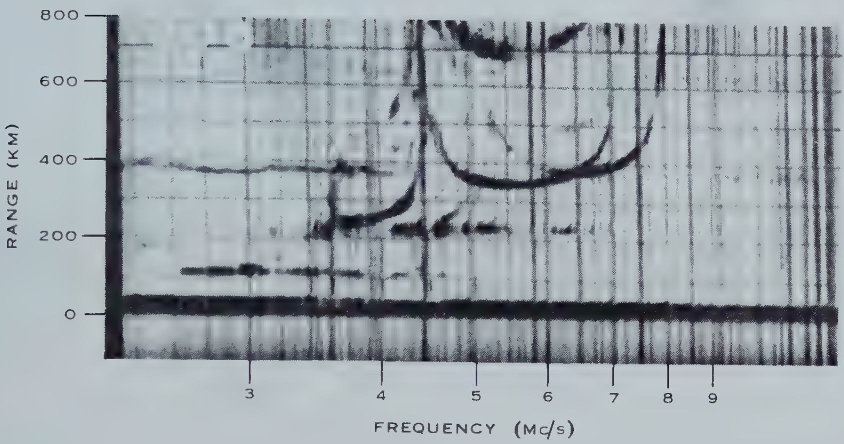


Fig. 1.—*P′f* record showing $f_mE_s > fE_s$.

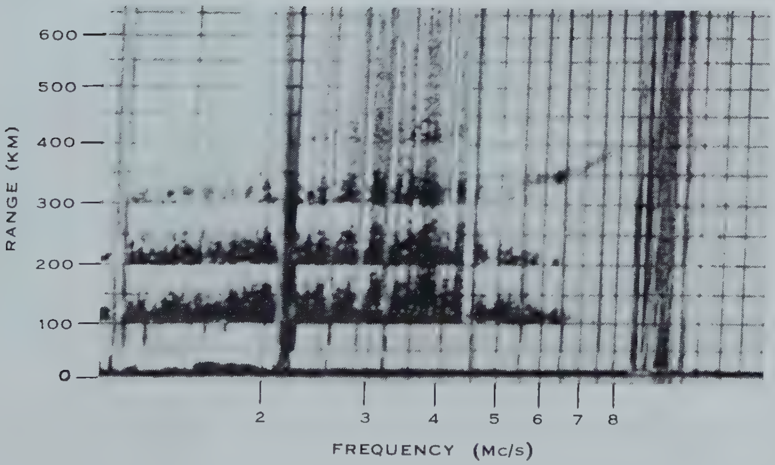


Fig. 2.—*P′f* record showing range spreading (index 4).

VII. Range Spreading

On many records the E_s trace shows the phenomenon of " range spreading ", that is, the returned echo is elongated to a much greater extent than can be accounted for by the limited bandwidth of the receiver (Plate 2, Fig. 2). As a rough guide to the extent of this phenomenon successive indices (starting from 1) have been allotted to traces which appear to be twice, three times, four times, etc. the normal trace width. Occasional traces with index 6 (i.e. ~140 km wide) have been observed. Figure 13 shows the mean value of this range-spreading index for the months of December and June. There is no evidence of solar cycle variation of this phenomenon. The marked peak and sudden drop after 1800 hr in June is evident in each year of the analysis. Range spreading occurs most frequently at times when f_bE_s is small, that is, when E_{sc} occurs, and this indicates that the E_{sc} region is very irregular in nature.

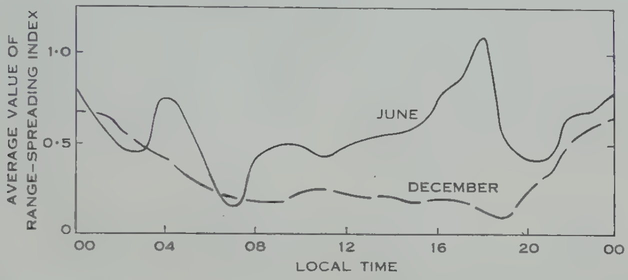


Fig. 13.—Summer and winter variation of range-spreading index.

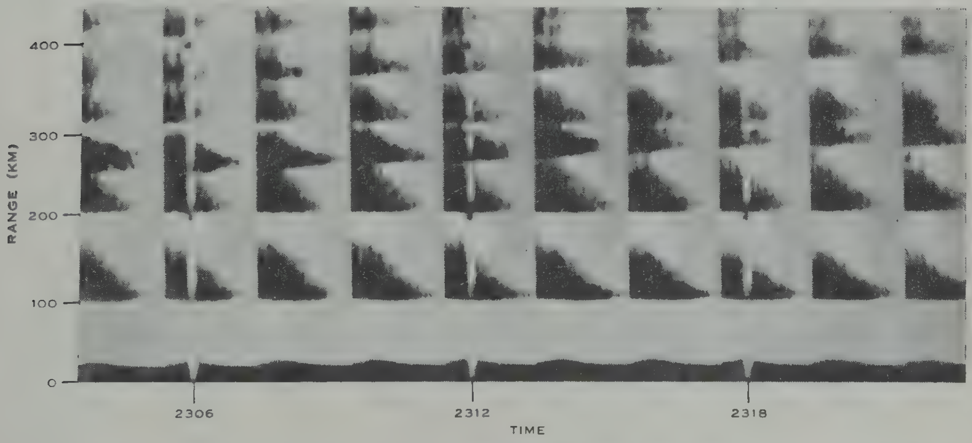


Fig. 14.—A typical swept-gain record at 2·28 Mc/s, showing $1E_s$, $2E_s$, $1F$, $3E_s$, $1F+1E_s$, and $4E_s$ echoes.

Range spreading observed at 2·28 Mc/s has been further examined by means of the swept-gain technique. A typical record is shown in Figure 14. Webster (personal communication) has shown that, provided there is no specular component to the echo, a measurement of the slope of the swept-gain patches gives a direct measure of the index n in the $\cos^n \theta$ polar diagram of Briggs and Phillips (1950). The results are more conveniently expressed in terms of the half-amplitude angle, θ_0, of the polar diagram of scattered radiation received. A histogram of the computed values of θ_0 for one typical night is given in Figure 15.

Direction-finding measurements at 2·28 Mc/s can occasionally be made on an isolated part of a range-spread echo. Such echoes are invariably found to be non-zenithal. Ample evidence is found of the horizontal movement of patches of ionization, and these movements are to be the subject of another paper from this laboratory. Such movements can sometimes be detected on $P'f$ records.

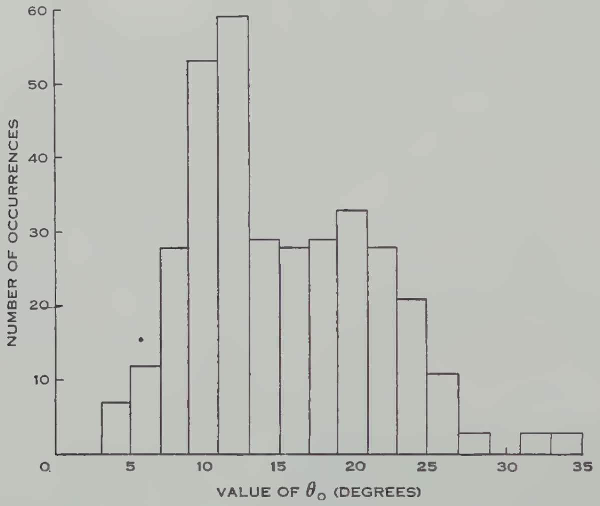


Fig. 15.—Values of θ_0 as determined from swept-gain records for the night November 16–17, 1954.

Measurements taken at 3·84 Mc/s and 5·8 Mc/s often show patches of sequential E_s drifting overhead ; there is a major distinction between such drifting clouds and those commonly seen at 2·28 Mc/s, which are nearly all due to patches of E_{sc}. E_{sc} clouds can be observed when still a considerable distance from the station and appear as " saucer-shaped " records on the $P't$ film ; E_{ss} clouds, on the other hand, show very little, if any, change of range as the cloud passes over. Phase-path measurements show that such E_{ss} clouds can only be seen when almost directly overhead.

VIII. Lateral Extent of E_s Patches

McNicol and Gipps (1951) have found that sequential E_s shows uniformity over wide areas. Sequences are recorded at the same times at recorders placed roughly on the same longitude, namely, Townsville (lat. 19·2 °S.), Brisbane (lat. 27·5 °S.), and Canberra (lat. 35·3 °S.). It was found, however, that the more southerly the station the less clear cut the sequence, the later the start, and the greater the tendency for decay to intervene before the sequence is complete.

Oblique incidence propagation of 73 Mc/s transmission from Melbourne to Brisbane, a distance of 1350 km, has often been reported* in summer. These reports have always occurred on days when fE_s at Brisbane was rather higher than normal, and hence when the E_s further south was also likely to be capable of reflecting high frequencies.

Constant-height type E_s, however, very often shows only limited correlation on fixed frequency records taken at stations separated by only 100 km. The average diurnal occurrence of E_{sc} is the same at each station but the spatial correlation is small. Records of slightly oblique reflections between pairs of these stations still show only limited correlation. Observations of isolated "clouds" which produce blanketing of the F region, indicate cloud sizes of the order of 10 km.

IX. Reflection Coefficients and Solar Cycle Control

On many occasions the reflection coefficient of the E_s layer is such that more than one reflection may be observed. On some occasions up to 20 reflections† have been recorded using a fixed frequency recorder operating on 2·28 Mc/s, and the occurrence of six or more reflections on the $P'f$ records is quite common. Diurnal variations of the average number of multiple echoes on $P'f$ records for summer and winter are shown in Figure 16. The day-time values of the effective reflection coefficient‡ is, of course, reduced by the presence of non-deviative absorption occurring below the E region level. But for this absorption, the number of multiples in December day-time would very probably be greater than in December night-time. This would indicate a higher reflection coefficient

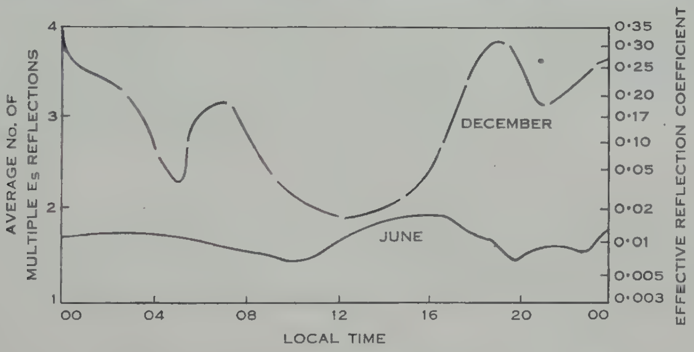


Fig. 16.—Multiple reflections and effective reflection coefficient for summer and winter.

---

* The author is indebted to the Radio Officer of the Queensland Ambulance Transport Brigade for these reports.

† Such a high number of multiple reflections is indicative of focusing effects as discussed by Baird (1954) and Schrag (1955).

‡ The effective reflection coefficient is found by comparing $P'f$ records with simultaneous swept-gain records at 2·28 Mc/s; the effective reflection coefficient may be found directly from the latter.

for E_{ss} than for E_{sc}. One might expect some changes in reflection coefficient with changes in critical frequency. Such changes, if they exist, are not sufficient to materially affect the number of multiple reflections; no kinks are observed in the curve for December corresponding to times of maximum critical frequency.

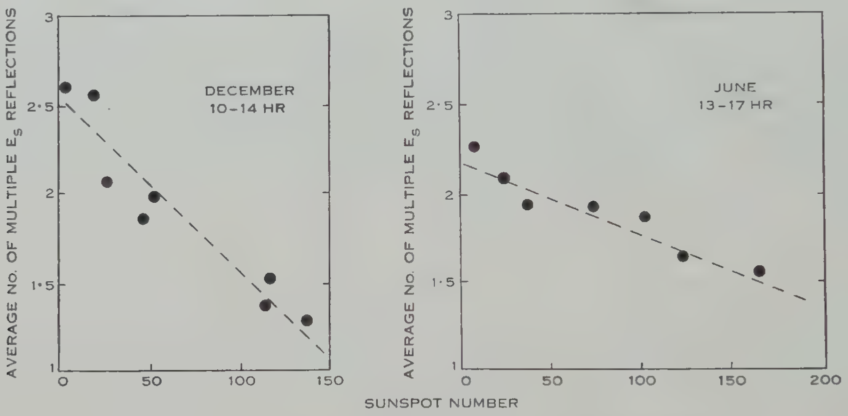


Fig. 17.—Sunspot cycle variation in E_s multiple reflections.

It thus appears likely that the reflection coefficient of E_{ss} is fairly constant throughout the daylight hours.

D region absorption is particularly evident in December. Figure 17 shows the variation in the average number of multiple reflections during the period 1000–1400 hr in December and 1300–1700 hr in June, as a function of sunspot number. Figure 18 shows the number of recordings of E_s echoes during 1000–1400 hr in December as a function of sunspot number. Both of these graphs indicate the strong dependence of the effective reflection coefficient on solar ionization. Since this is the *only* property associated with E_s reflections

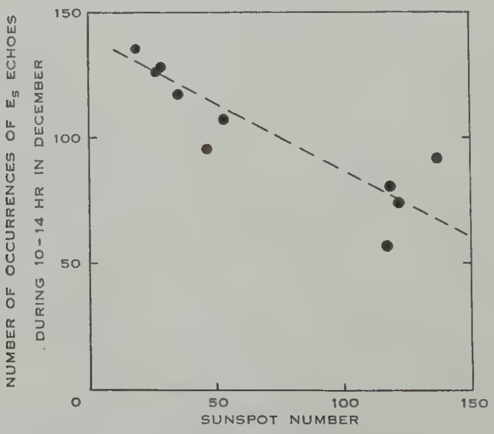


Fig. 18.—Sunspot cycle variation in E_s echoes recorded.

which shows any solar cycle dependence, it is a fair assumption that the majority of the solar cycle change in the effective reflection coefficient is due to a change in absorption occurring in the D region.

X. Correlation with Other Phenomena

(*a*) *Meteor Correlation*

The earlier paper by McNicol and Gipps (1951) pointed out that there is very little correlation between E_s at Brisbane and the occurrence of meteoric dust. Naismith (1954) has reported a " meteoric E " layer in England which is responsible for the majority of Slough records scaled previously as E_s. The diurnal distributions of the critical frequency and occurrence of this layer are entirely different from those for E_s at Brisbane. Occasional weak echoes are seen on the fixed frequency records below the normal E_s level. It is possible that these may correspond to such a meteoric E layer, only being observed when the reflection coefficient is sufficiently high.

(*b*) *Thunderstorm Correlation*

According to one theory of E_s formation, thunderstorms may provide a means of producing new ionization in the E region. Mitra and Kundu (1954) have reported an increase in E_s ionization coincident with the onset of squall-type storms in Bengal in the pre-monsoonal months. Reliable records of thunderstorm occurrences and paths are available in Brisbane, and a search through the records showed no E_s-thunderstorm correlation, even in the pre-monsoonal months of October and November.

XI. Conclusion

It has been found possible, on a statistical basis, to achieve a fairly detailed separation of the occurrence of the two types of E_s at and below 120 km. The chief instrument in this separation is the recorded value of the *blanketing frequency*.

(*a*) *Sequential* $\mathrm{E_s}$

E_{ss} is predominant in summer, there being normally two sequences starting at about 0530 hr and 1300 hr. In both morning and afternoon the time of maximum intensity of E_{ss} ionization occurs about 4–5 hr after the breakaway from E_2, that is, about 4 hr after the ionization is at its maximum height.

Sequential E_s occurs with decreasing regularity and strength until September, after which there is a rapid rise to the peak in December.

Many workers have discussed the types of ionization distribution which could lead to the conditions holding in the early stages of sequential E_s (Best, Farmer, and Ratcliffe 1938 ; Appleton, Naismith, and Ingram 1939 ; Rawer 1939*a*, 1939*b*, 1940, 1949 ; Briggs 1951*a*, 1951*b* ; Whale 1951 ; Chatterjee 1953). Briggs (1951*b*) has shown that when the frequency overlap (or frequency range of partial penetration and partial reflection) is small, then the ionization is in the form of a thin layer of high ionization density, with absorption occurring near the critical frequency. The increase in the frequency overlap from zero at the beginning of a sequence, to 1 Mc/s or more in the later stages thus suggests a very substantial thinning down of the ionized region. Martyn (1953) has shown that a mass of excess ionization which moves in a region where Hall conductivity is appreciable will accrete ionization on one side at the expense of the other, the ionization building up into a thin shell on the leading edge of the moving mass. If the stratification and breakaway of ionization from around the 200 km level were due to solar tidal motion in the upper atmosphere, then E_{ss} formation could be satisfactorily explained on that basis.

The increased number of E_2 records obtained in winter and spring suggests that the tidal forces causing breakaway of ionization from the E_2 region are stronger in summer and autumn than in the other two seasons.

The facts that (i) the maximum in fE_s lags behind the maximum virtual height by 4–5 hr, and (ii) the lunar tidal analyses of Matsushita (1953) show a lag of fE_s on $h'E_s$ of about 2 hr, indicate some measure of support for the hypothesis of tidal drift of ionization, as the source of E_{ss}—accretion of ionization will not be well marked until the region has been moving downwards for an appreciable time. It is rather difficult, however, to account for the lack of increase in critical frequency with an upward drift, which exists, for example, between the morning and afternoon downward drifts in summer months.

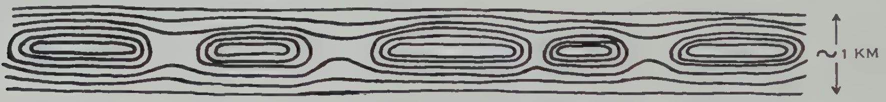


Fig. 19.—Suggested ionization contours in the latter stages of sequential E_s.

The coincidence of sequential movement at the three widely spaced stations along roughly the same latitude shows that the E_2 stratification and subsequent breakaway are indications of the widespread effect of some type of solar control. It is obviously not a direct solar control since there is often decay and re-formation of E_2 after midday in summer, followed by another sequence. The most plausible explanation of such double sequences lies in solar tidal drifts of ionization.

In the later stages of the sequence (below about 120 km) the range of partial transmission extends to more than 1 Mc/s and the $P'f$ record may take on the patchy appearance often associated with E_{sc}. It is probable that at this stage the " thin layer " is breaking up into discrete regions or " clouds " of ionization, remaining fairly thin, but with the gaps between the clouds still retaining a reasonably high electron density and hence capable of producing blanketing to quite high frequencies (Fig. 19). Such thin disk-like clouds would be capable of giving substantial reflection only when almost exactly overhead. In the final stages the overall ionization density decreases so that both fE_s and f_bE_s decrease together.

(*b*) *Constant-height Type* E$_s$

The " patchy " appearance often associated with E_{sc} traces is indicative of a considerable amount of frequency fading. Briggs (1951*a*) has found such frequency fading to be associated with a region capable of giving scattered reflections at a considerable angle to the vertical; such reflections are not found for the sequential type of E_s.

There is ample evidence of non-zenithal reflections from E_s clouds and the nature of this type of E_s is not much in doubt. E_{sc} consists of ionization in which is imbedded patches or clouds of enhanced ionization. This background of ionization may, or may not, be observed.

The oblique reflections obtainable from E_{sc} ionization are responsible for the range spreading referred to in Section VII. The late afternoon winter peak in range spreading indicates that immediately before decay the region tends to become very patchy or rough and can thus reflect strongly from non-zenithal points.

Experiment has eliminated three theories for the origin of E_{sc} at Brisbane. Booker's (1950) scattering theory is untenable since fE_s is independent of recorder sensitivity. The negative correlation of E_s with meteor and thunderstorm occurrence eliminates these two as causes of E_{sc}.

The true cause of E_{sc} ionization is not known with any certainty. The most probable explanation lies again in tidal drifts due to solar forces. The seasonal shift in the times of maximum occurrence of high critical frequency (keeping 4–5 hr behind maximum $h'E_s$) cannot well be explained on any other hypothesis.

XII. Acknowledgments

The author wishes to thank Professor H. C. Webster and Mr. R. W. E. McNicol for their continued guidance and helpful discussion. This work has been carried out at the University of Queensland as part of the general programme of the Radio Research Board, C.S.I.R.O., and is published by permission of the Board.

XIII. References

Appleton, E. V., Naismith, R., and Ingram, L. J. (1939).—*Proc. Phys. Soc. Lond.* **51**: 81.
Baird, K. (1954).—*Aust. J. Phys.* **7**: 165.
Best, J. E., Farmer, F. T., and Ratcliffe, J. A. (1938).—*Proc. Roy. Soc.* A **164**: 96.
Booker, H. G. (1950).—Proc. Pennsylvania Conf. Ionospheric Physics A, p. 171.
Briggs, B. H. (1951*a*).—*Proc. Phys. Soc. Lond.* B **64**: 255.
Briggs, B. H. (1951*b*).—*J. Atmos. Terr. Phys.* **1**: 345.
Briggs, B. H., and Phillips, G. J. (1950).—*Proc. Phys. Soc. Lond.* B **63**: 907.
Brown, J. N., and Watts, J. M. (1950).—*J. Geophys. Res.* **55**: 179.
Chatterjee, B. (1953).—*J. Atmos. Terr. Phys.* **3**: 229.
Dieminger, W. (1951).—*Proc. Phys. Soc. Lond.* B **64**: 142.
Helliwell, R. A. (1952).—Paper presented at U.R.S.I. meeting in Sydney, June 1952.
Lindquist, R. (1953).—*J. Atmos. Terr. Phys.* **4**: 10.
McNicol, R. W. E., and Gipps, G. de V. (1951).—*J. Geophys. Res.* **56**: 17.
McNicol, R. W. E., Webster, H. C., and Bowman, G. G. (1956).—*Aust. J. Phys.* **9**: 247.
Martyn, D. F. (1953).—*Phil. Trans.* A **246**: 306.
Matsushita, S. (1953).—*Rep. Ionosphere Res. Japan* **8**: 45.
Mitra, S. K., and Kundu, M. R. (1954).—*Nature* **174**: 798.
Naismith, R. (1954).—*J. Atmos. Terr. Phys.* **5**: 73.

Rawer, K. (1939*a*).—*Hochfrequenztech. u. Electroakust.* **53** : 150.
Rawer, K. (1939*b*).—*Ann. Phys.* **35** : 385.
Rawer, K. (1940).—*Naturwissenschaften* **28** : 577.
Rawer, K. (1949).—*Nature* **163** : 528.
Schrag, R. L. (1955).—" The Physics of the Ionosphere." p. 202. (Physical Society : London.)
Thomas, J. A., and McNicol, R. W. E. (1955).—*Proc. Instn. Elect. Engrs.* **102 B** : 793.
Thomas, J. A., and Svenson, A. C. (1955).—*Aust. J. Phys.* **8** : 554.
Thomas, J. A., Svenson, A. C., and Brown, H. E. (1956).—*Aust. J. Phys.* **9** : 159.
Watts, J. M., and Brown, J. N. (1951).—*J. Geophys. Res.* **56** : 403.
Whale, H. A. (1951).—*J. Atmos. Terr. Phys.* **1** : 233.

A STUDY OF "SPREAD–F" IONOSPHERIC ECHOES AT NIGHT AT BRISBANE

I. RANGE SPREADING (EXPERIMENTAL)

By R. W. E. McNicol,* H. C. Webster,* and G. G. Bowman*

[*Manuscript received January* 31, 1956]

Summary

At frequencies well below the critical frequency, satellite echoes sometimes accompany the night-time F_2 echo, sometimes clearly separated, sometimes overlapping. In an investigation of these range multiplets, in addition to routine $P'f$ sounding records, continuous virtual range measurements at fixed frequency (at stations of various separations), and measurements of mean intensities, phase-path changes, and directions of arrival, have been carried out.

From a study of the results, certain simple relationships emerge, as first approximations, between the various quantities measured, namely: virtual ranges (group paths), phase-path change, zenith angle, azimuth, intensity, and the time. A parameter, having the dimensions of velocity, appearing in these relationships, has values of which 80 per cent. lie in the range 240 ± 140 km/hr, with directions of which 80 per cent. lie in the range $290 \pm 60°$. The mean duration of a satellite is 50 min. Satellites occur more frequently in winter than in summer and show a somewhat complex diurnal variation, which is described in detail, and which tends to recur from year to year. It is associated with the general night-time changes in equivalent height of the layer. Strongly reflecting E_s shows an inverse diurnal variation. No correlation with geomagnetic disturbance indices could be found.

I. Introduction

No one who examines regularly the night-time records of a swept-frequency ($P'f$) ionospheric sounding equipment operating at medium latitudes can fail to remark the diffuse character frequently exhibited by the F reflections at night. Although in routine reports from Brisbane (lat. 27° 30′ S., long. 153° 00′ E., geomag. lat. 35° 06′ S.) such occurrences are simply noted as "spread or diffuse echo", the phenomenon assumes, in fact, a wide variety of forms. Thus there may be only a broadening of the echo traces, of roughly the same extent, at all frequencies, giving a broad band on the record (narrowing as it sweeps upwards), as in Plate 1, Figure 1. On the other hand, the broadening may occur chiefly near the penetration frequency axis where the traces sweep upwards, as in Plate 1, Figure 2. The two effects can be present together and both can attain considerable magnitude, as in Plate 1, Figure 3, so that on the one hand, much of the range of virtual heights between the F and $2F$ traces is

---

* Physics Department, University of Queensland, Brisbane.

filled in with reflections, and on the other hand, a range of frequency of several megacycles per second is covered by the broadened upward sweep.

As reported by Gipps, Gipps, and Venton (1948), the broad horizontal band at lower frequencies is sometimes replaced by two or more distinct traces running parallel to each other, with constant separation in group path over most of the frequency band, as shown in Plate 2, Figure 1.* These traces usually, but not always, give the same penetration frequencies. Plate 2, Figure 3, shows a case in which the trace of greater group path gives higher penetration frequencies. In such cases the penetration frequencies usually change little, if at all, between successive $P'f$ records (taken at 10-min intervals).

Again, the horizontally broadened trace near the penetration frequency may be replaced by a few distinct upward-sweeping traces, as illustrated in Plate 2, Figure 2. Detailed examination excludes the possibility that any of these are due to the so-called Z reflection.

It is believed, in the light of all the available evidence, that the broadening of the traces in the $P'f$ records at frequencies well below the critical frequency is due to the presence of a large number of individual traces, which are not resolved.

It is convenient to refer to such occurrences as " multiplets ". The occurrence of multiple traces, with different group paths, at low frequencies will be termed a " range multiplet " (cf. Plate 1, Fig. 1) ; where the distinction lies chiefly in the penetration frequencies it will be termed a " penetration frequency multiplet " (cf. Plate 1, Fig. 2). The latter will be the subject of another paper from this laboratory.

The term " equivalent range " (R') rather than " equivalent height " is used for the semi-group path, because, as will be shown later, there is evidence that the echoes of range greater than the minimum are due to non-vertical reflection. When there is only one echo of increased range, or when only one such echo is clearly separated from its neighbours, the phenomenon will be described as a " range doublet " (cf. Plate 2, Fig. 1).

As pointed out by Gipps, Gipps, and Venton (1948), the separations in range of the components of a range doublet usually show a progressive change from frame to frame of the swept-frequency records (taken at 10-min intervals). Since the separations are independent of frequency, complicated changes can be followed much more readily by continuously recording echoes from a transmitter of suitably chosen fixed frequency. Such records form the main study of this paper. With a fixed frequency it is also practicable to carry out a number of ancillary measurements on the returning signals, as discussed below.

While this work was in progress, a paper by Uyeda and Ogata (1954) appeared which reports similar results.

Discussion of the physical significance of the observations is given in the second paper of this series.

---

* $P'f$ records of rather similar appearance to this may be produced by double-hop reflection, first from the F region, then from the E ; it is assumed that these are excluded.

II. Fixed-frequency Equipment

(*a*) *Virtual Range; Fixed-gain Records*

Three complete ionospheric sounders, operating mostly on 2·28 Mc/s were used, located, for the most part, at Brisbane, Buderim (95 km north of Brisbane), and Toowoomba (95 km west of Brisbane).

The transmitter power was 500 W, pulse duration 60 μsec, 50 pulses/sec. The receiver bandwidth was such that received pulses were resolved, in favourable cases, if their group paths differed by more than 25 km. The transmitter and receiver both used half-wave dipole aerials 50 ft above ground. The gain of the receiver was adjusted so that echoes down to 60 dB weaker than the strongest ionospheric echoes were clearly recorded.

Output pulses from the receiver were applied to a 6 in. cathode-ray tube in such a way as to black out (or, in some cases, brighten) the fluorescent spot. The tube was supplied with a linear time-base, the speed and time of starting of which were adjusted according to the particular group of echoes to be recorded. To improve resolution in the records two or more tubes were sometimes used, with different (expanded) portions of the time-base displayed on each. " Range marks ", generated by a stable oscillator, were applied to the tube for short periods at regular intervals. (Continuous application might have resulted in obscuration of some echo traces.) These brief applications of range marks also served as time marks. Small neon glow tubes, mounted alongside the cathode-ray tube and operated by a clockwork-driven contactor, provided a code for distinguishing the hours. Photographic records were made on 35-mm film using a camera with $f/4\cdot5$ aperture, with a film speed of 5 in/hr.

Simultaneous observations using all three stations have been made on more than 100 nights; and on many more nights using only one or two of the stations. A few observations were also made at Goondiwindi (about 300 km south-west of Brisbane). The transmitters were coded for identification.

Records were also made at Brisbane of echoes originating from the Buderim and Toowoomba transmitters. These will be referred to as TR' records, to distinguish them from those for which transmitter and receiver occupy the same site, which will be designated TR. Phase jitter (usually around ±50 μsec) led to a spurious broadening of the traces received; nevertheless, the records contain useful information. An example of a TR record is given in Plate 3, Figure 1, and of a TR' record in Plate 3, Figure 2.

(*b*) *Virtual Range; Swept-gain Records*

Records have been made at Brisbane using a receiver in which the gain was controlled by a motorized potential divider rotating once in 2 min. The gain of the receiver fell in steps of 2 dB, then rose suddenly, by 80 dB, to maximum gain. At minimum gain no echoes are recorded. The swept-gain records thus consist of regularly broken traces, the length of each segment being a measure of the relative intensity (in dB) of the particular echo (cf. Plate 3, Fig. 3). Rapid fading tends to be averaged out, but slow fading can affect the measurements unless an average is taken over several segments.

Swept-gain records tend to show improved resolution, for, when the receiver gain reaches a value such that a particular echo only just records, only the peak of the echo shows. Thus towards the ends of the segments in the swept-gain records the traces narrow; this often permits the separation of closely spaced traces which, in the fixed-gain record, show as a single broad trace. Falconer (personal communication, 1955) has found that, even when separation is not possible in individual segments, an averaging process over several segments can reveal the major components of the broadened trace.

(c) "*Phase-path*" *Records*

Equipments for measuring changes in the phase path traversed by ionospheric echoes have been described by Findlay (1951) and Jones (1953). The equipment used in Brisbane was a simplification of that of Jones, and required only small modifications of the standard receiver-display combination. An external oscillator, tuned to a frequency about 40 kc/s above the intermediate frequency of the receiver, was switched on once each cycle by a rectangular voltage pulse, starting simultaneously with the modulator pulse for the transmitter. The initial phase of the oscillations was fixed by the injection into the oscillator circuit of an intermediate-frequency pulse derived, in the receiver, from the "ground" pulse of the nearby transmitter. A suitable fraction of the oscillator output was mixed in a non-linear circuit with the full intermediate-frequency output of the receiver, and the 40 kc/s component of the mixture was filtered out and rectified. This video output was fed to a display system similar to that described in Section II (*a*). A film speed of 15 in/hr was found necessary to record accurately the rapid phase-path changes which sometimes occurred. Plate 4 shows portion of a typical record.

(*d*) *Direction-of-arrival Records*

Special equipment has been installed at Brisbane for the determination of the direction of arrival (azimuth and altitude) of ionospheric echoes at 2·28 Mc/s. This is described elsewhere (Thomas and McNicol 1955). Automatic records provide simultaneous information concerning a whole group of echoes. The method fails in the case of echoes which overlap each other, as in a multiplet of small range separation.

III. EXPERIMENTAL RESULTS

(*a*) *Satellites and their General Characteristics*

The frequency of 2·28 Mc/s is convenient for the investigation of range multiplets, since the *o*- and *x*-rays are rarely separated at this frequency. Range multiplets show on the *R't* records as a multiplicity of traces, as exemplified in Figure 1, and Plate 3. It is convenient to distinguish the lowest trace, which is usually continuous over long periods (sometimes all night) from the upper traces, which show continuity for shorter periods, sometimes only a few minutes (cf. Fig. 1). The former is therefore termed the "main trace" and the latter are termed "satellites".

The range separation between satellite and main trace usually increases or decreases considerably, and systematically, during the lifetime of the satellite; a range doublet or multiplet is therefore classified as divergent, convergent, parallel, or mixed, according to the behaviour of the satellites constituting it. When the successive points of junction between satellites and main trace are spaced at intervals less than about 2 min, the satellites are regarded as belonging to one and the same multiplet; this criterion is, of course, quite arbitrary. An example of a convergent range doublet is given in Figure 2 and of a mixed range multiplet in Plate 3. The relative frequencies of occurrence of parallel, convergent, divergent, and mixed range doublets are roughly in the ratio of 1 : 8 : 20 : 6.

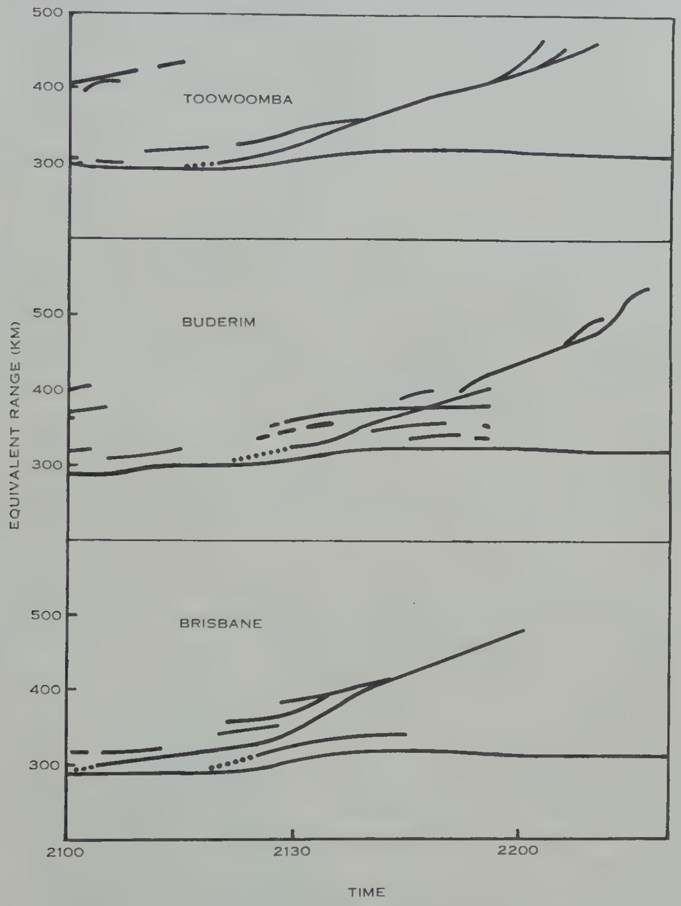


Fig. 1.—Divergent range multiplet recorded simultaneously at three separated stations. Portion of $R't$ records for Toowoomba, Buderim, and Brisbane, July 22, 1950, replotted from readings. Some of the traces on the original records were broader than is shown and closely adjacent traces cannot be resolved from each other; the dotted portions are thus conjectural.

It has been found that satellite traces usually, but not always, show a slight upward curvature. They often satisfy approximately the semi-empirical hyperbolic relationship

$$\sqrt{(R'^2-h'^2)}=k(t-t_0),$$

where h' represents the *lowest* range observed during the lifetime of the satellite,* and k and t_0 are constants (k may be negative). As an example, Figure 3 shows a plot of $x(=\sqrt{(R'^2-h'^2)})$ versus t in a favourable case. The slope dx/dt $(=k)$, which has the dimensions of speed, has been determined† from a large number of satellites and the distribution of speeds is shown in Figure 4. The mean value is 63 m/sec (230 km/hr). In Figure 5 the results for convergent and divergent satellites are shown separately. In view of the small number of cases, it is questionable whether the difference in distribution is significant.

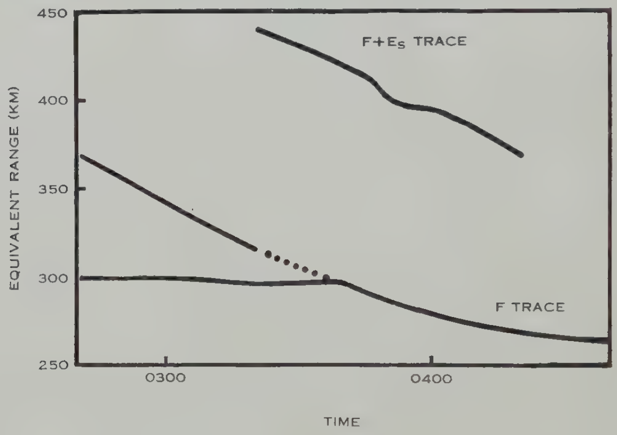


Fig. 2.—Convergent range doublet. $R't$ record for Brisbane, May 26, 1952. Replotted from readings. The original traces are considerably broader than the diagram would indicate and the final junction of the satellite is thus obscured ; a possible extrapolation is indicated by the dotted line. Note the parallelism between the $(F+E_s)$ trace (reflection first from F, then ground, then E_s) and the satellite.

Some satellite traces (in ordinary fixed-frequency records) show wide gaps and some resume, after a gap, with a trace corresponding to a slightly different value of t_0. These complications introduce a measure of arbitrariness into the interpretation of the records.

It is clearly of some interest to know whether the changes in R' arise from change in actual distance of the reflecting point, or from group-retardation effects. The horizontality of the swept-frequency traces at low frequencies makes the latter hypothesis improbable, but the point was more definitely checked by comparing changes in phase path with the simultaneous changes in group path. Figure 6 shows the results of such a comparison. Except when special conditions operated, as discussed below, discrepancies between group and phase path changes exceeding 10 per cent. were never clearly demonstrated. It should be pointed out, however, that the fringes in phase-path records were sometimes confused. In certain of these confused cases there was evidence of overlapping and intersecting satellites. (Nevertheless, complicated records could sometimes be disentangled more readily from the phase-path than from the ordinary records.)

---

* In reading h' it is assumed that a given trace retains its identity through any intersection or junction as long as there is no discontinuity in its slope. Thus the " lowest " range of a satellite will usually be read off a portion of the trace produced beyond the intersection with the main trace.

† The authors are indebted to Mr. D. Lamb for analysing these records.

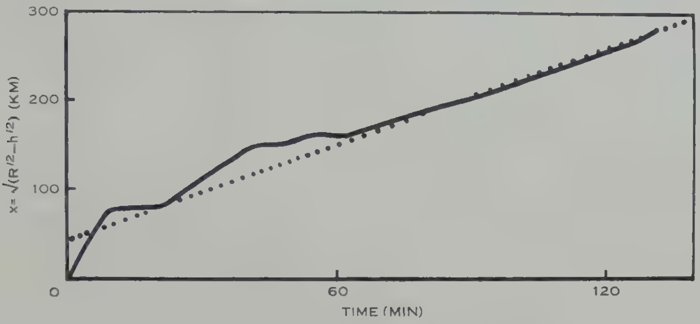


Fig. 3.—Plot of $x(=\sqrt{(R'^2—h'^2)})$ versus t. (Brisbane, October 16, 1954.) The full curve joins the points plotted, the broken curve is the best-fitting straight line. Note that since $\mathrm{d}x/\mathrm{d}R'$ becomes infinite as x approaches zero, the lower part of the curve is subject to great error.

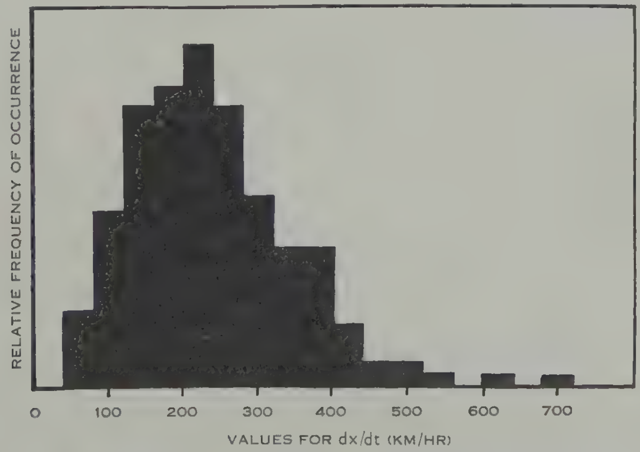


Fig. 4.—Distribution of values of $\mathrm{d}x/\mathrm{d}t$ for all satellites in 1952.

An important characteristic of the majority of range multiplets is the close equality between the penetration frequencies (as indicated by the upward sweep of the curves on swept-frequency records) of satellites and main trace. A difference between satellite and main trace critical frequencies was detected on less than 10 per cent. of occasions. In such cases, the satellite critical frequency can be either lower or higher than that of the main trace; the former condition seems rather more prevalent, but this impression may arise because it is experimentally easier to detect. Sometimes the frequency difference gradually disappears during the currency of the satellite. Plate 2, Figure 3, indicates one of the rare cases when the satellite has the higher critical frequency; it may be significant that the corresponding $2 \cdot 28$ Mc/s $R't$ trace has an unusual speckled appearance.

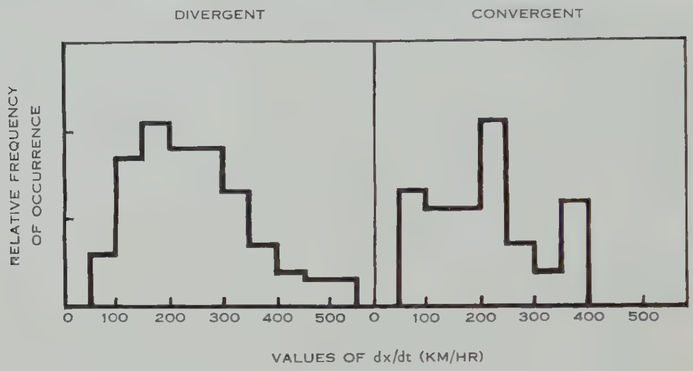


Fig. 5.—Distributions of values of dx/dt for 110 divergent and 79 convergent satellites separately.

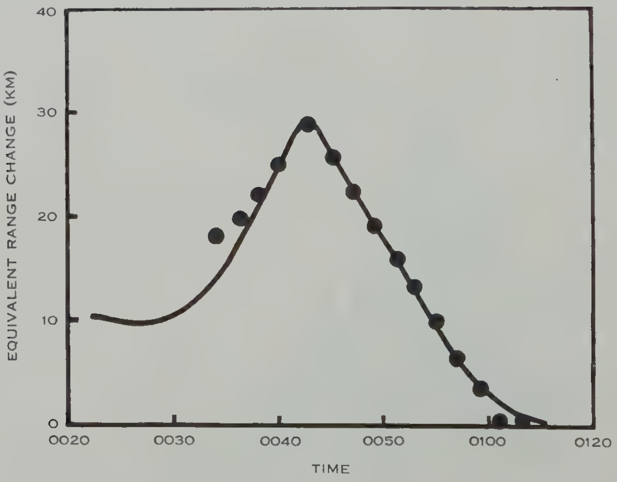


Fig. 6.—Comparison of group-path and phase-path changes. Changes of group path (full line) and phase path (dots) for lowest ("main") trace of a mixed range multiplet (June 23, 1955, at 0040 hr). The two plots have been adjusted to coincide at the peak.

It is perhaps worthy of note that the peculiar satellite traces (characteristic of oblique-incidence propagation) reported by Meek (1952) and others in Arctic recordings are never observed in Brisbane.

Determinations of directions of arrival show that, for the main trace, the echoes nearly always arrive vertically (within limits of experimental error) while the satellite echoes arrive at an angle with the vertical. It is found that in nearly all cases, the zenithal angle θ satisfies approximately the relationship

$$\sec\theta = R'/h'.$$

Table 1 shows the results of a comparison of values of θ deduced from the expression and values determined directly. Departures are of the same order of magnitude as the estimated experimental errors.

TABLE 1

CALCULATED AND OBSERVED ZENITH ANGLES FOR PARTICULAR SATELLITES*

Time	Zenith Angle, θ	
	Calculated (deg)	Observed (deg)
2309	43	46
2330	36	32
2351	41	44
0106	25	26
2206	38	36
2236	24	20
2330	30	32
0000	52	54
0151	16	15
0218	40	44
0203	38	37
0045	34	35
1957	27	27
0254	53	54
1930	24	28
2218	48	48
2100	14	0
2245	40	53
2257	41·5	39
2303	43·5	42
2227	32	33
2303	29	26
2315	31	22 ?
2321	34	34
2345	37	37
2351	40	40

* The authors are indebted to Mr. W. Woolcock for many of these values.

(*b*) *The Junction or Intersection of a Satellite with the Main Trace*

The general appearance of many mixed multiplets, especially those where there are only a small number of satellites, strongly suggests the crossing-over of two sets of traces, i.e. the lower set after intersection is continuous with the upper set before intersection and vice versa (cf. Plate 3). In such cases it is assumed that the roles of main trace and satellite are exchanged at the intersection. Measurements of the azimuths of arrival of the echo giving the lowest trace confirm this view. On several occasions it was possible to determine the azimuths; it was found that an abrupt 180° change occurred at the intersection. Furthermore, an $R't$ plot, using the lower trace on both sides of the intersection, shows a characteristic change of slope at that point. This is brought out in Figure 7, which is an average taken over a number of cases, setting the time zero at the point of intersection. It will be noted that the average range at intersection is 10 km higher than at some time previously and subsequently.

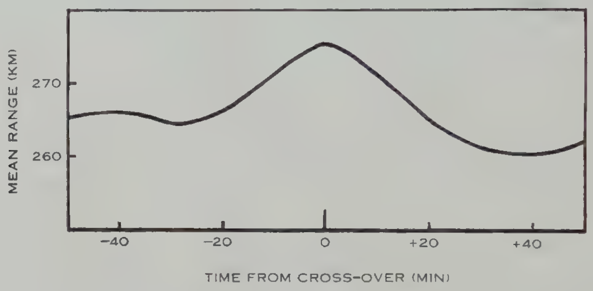


Fig. 7.—Average course of the lowest trace in the vicinity of the intersection of the traces in 54 mixed range multiplets.

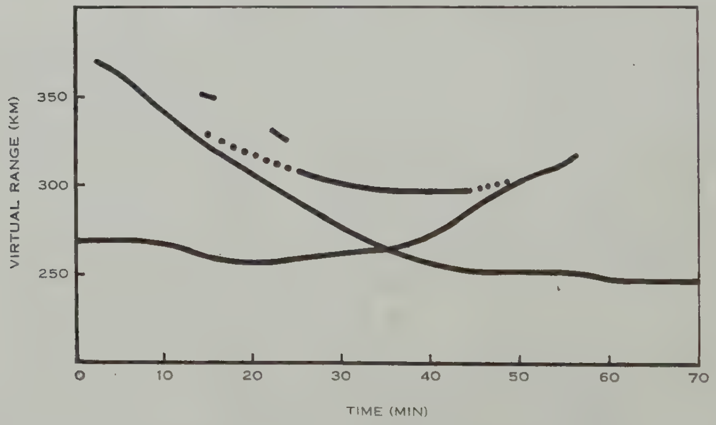


Fig. 8.—Mixed range multiplet showing traces of upper member $R't$ record for Brisbane, March 17, 1953, replotted from readings. The dotted portions are conjectural, since the lower trace was broader than the diagram suggests and these portions of the upper satellite, if present as indicated, would have been obscured.

A few range doublets of the mixed type show fragments of a third satellite forming a roughly parabolic curve connecting the ends of the other two satellites (cf. Fig. 8).

For the two major intersecting traces in a mixed range doublet, designated A and B, the value of $x_A - x_B$ is approximately constant. Values of this quantity for nine selected traces are plotted in Figure 9 against the corresponding values of dx/dt. There is no indication of any definite relationship between the two variables. The average value of $x_A - x_B$ for the cases plotted is about 280 km.

The durations (and intensities) of the satellites on the two sides of the intersection are often very different. We can therefore regard convergent and divergent range doublets as limiting cases of mixed doublets in which one satellite has completely disappeared. In support of this view it is noted that a satellite in a convergent or divergent doublet frequently (but not always) appears to be continuous with the main trace on the far side of the junction (cf. Fig. 2). Moreover, an $R't$ plot, again using the lower trace throughout, shows a change of slope at the (estimated) junction point in these cases also. Figure 10 illustrates the average behaviour for 20 convergent and 75 divergent doublets respectively.

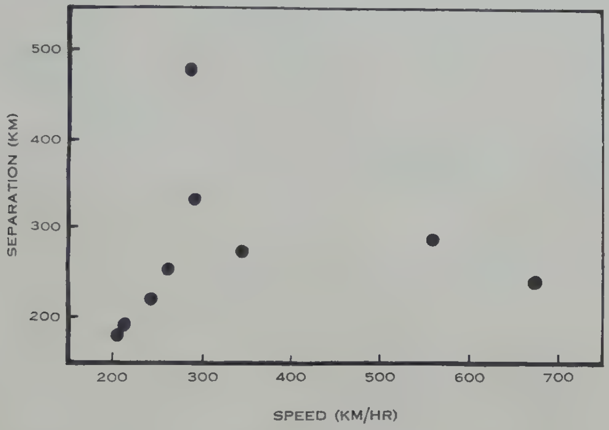


Fig. 9.—Plot of $(x_A - x_B)$ against dx/dt for mixed doublets.

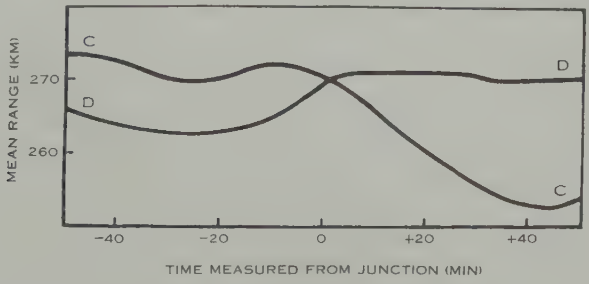


Fig. 10.—Average course of the lowest trace in the vicinity of the junction of the traces in 20 convergent doublets (curve C) and 75 divergent doublets (curve D).

While there are fairly wide differences between individual cases, the small increase in range of the main trace at the time of junction, for a divergent satellite, and small decrease for a convergent satellite is almost always noted.

Only a few satellites, and those mostly of short duration, fail to meet the main trace, if present, at some time during their existence. However, during periods when strongly reflecting E_s is present (indicated by multiple E_s reflections), so that all F echoes, on 2·28 Mc/s, are " blanketed " for long intervals, the record appears to consist only of satellite traces (cf. Fig. 11). Some satellites show a range decrease to a minimum, followed by an increase.

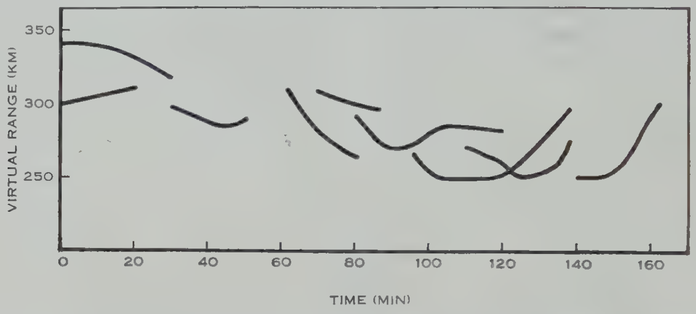


Fig 11.—Absence of continuous main F-region trace. Replot of portion of $R't$ record for Brisbane, Oct. 21–22, 1954. Strong E_s reflections were present but have not been plotted.

(c) R't *Records when Magneto-ionic Splitting is Apparent*

On a few occasions, because a higher working frequency (5·8 Mc/s) was used, or because the critical frequency was abnormally low, or group-retardation effects present, records of multiplets were obtained in which *o*- and *x*-rays

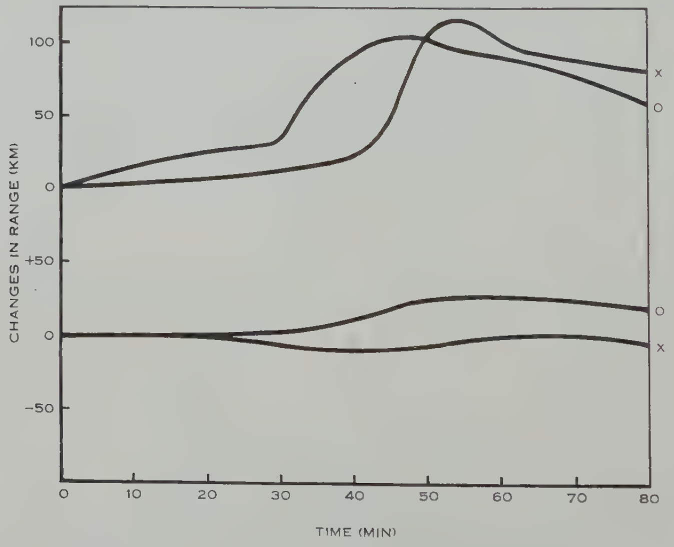


Fig. 12.—Comparison of group-path changes (upper curves) with phase-path changes (lower curves) during a rapid change involving separation of *o*- and *x*-rays, Mar. 5, 1955, starting at 0400 hr. Note that the phase path record for the *x*-ray was very weak and this curve is not wholly reliable.

produced clearly separated traces. (The possibility of mistaking a simple separation due to this cause for a range doublet was excluded by always checking against the swept-frequency records.) In many such cases, there was evidence of separation of *o* and *x* on the satellite, as well as on the main trace. As a general rule the satellite traces showed large gaps; correspondingly, it frequently happens that the satellite trace on swept-frequency records is incomplete, especially near the critical frequency.

Another phenomenon sometimes observed is a brief separation of *o*- and *x*-rays, accompanying a rapid increase in group path; this is termed a " lenticel " because of its appearance on the *R′t* records. In such cases, the group path increases much more rapidly than the phase path (cf. Fig. 12). Rapid group-path rises, not accompanied by separation of *o*- and *x*-rays, show similar, but less marked, discrepancies.

Lenticels are frequently accompanied by satellites, branching off where the rapid group-path rise ceases. Upon an occasion when a satellite was present prior to the lenticel, it showed a similar rise in group path and evidence of separation of *o*- and *x*-rays.

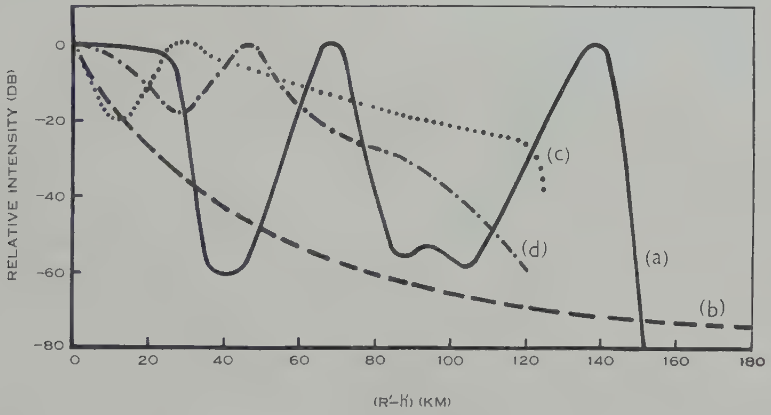


Fig. 13.—Changes of intensity (in dB) of satellite plotted against excess of range over minimum. Curve (*a*) represents a case of extreme fluctuation; curve (*b*) a case of unusually uniform change. Curves (*c*) and (*d*) represent more usual behaviour.

(*d*) *Intensities of Satellites*

When the separation of the satellite and main trace is barely sufficient to permit resolution, the intensities are usually equal, within a few decibels. As the satellite range increases, the echo intensity falls. Figure 13 shows the behaviour in some typical cases. The intensity changes are highly irregular, often showing successive maxima and minima, different by 40 dB or more. In Figure 14 intensity I (in dB) is plotted at intervals against range increment ($\Delta R'$) for a number of independent traces, taking the intensity (dB) and range increment zero at minimal range. (The sets of points have not been connected up.) The scatter diagram produced is crudely consistent with a linear relationship between I and $\Delta R'$, with a slope of $-0 \cdot 4$ dB/km. In spite of this general tendency to weaken with increasing range, some divergent satellites first appear at a range considerably above that of the main trace, and conversely some convergent satellites weaken markedly, or disappear, before they reach the main trace.

A cursory study was made of the relative intensities of satellites and the $(F+E_s)$ echo. Many multiplets, including the more complex ones, are unaccompanied by $(F+E_s)$. Where present, it is usually 15 dB or more weaker than the main F echo, and is weaker than the satellites lying between it and that echo.

(e) *Multiple-hop Satellites*

When a satellite appears on the one-hop F echo, in 20–30 per cent. of cases a satellite appears also on the two-hop, the separation between the respective satellites remaining constant throughout their duration. On a very few occasions, two $2F$ satellites have been observed, the second (weaker) satellite having a separation from its main trace which is around two-thirds of the separation of the other satellite (cf. Fig. 15). The second satellite may start later than the first. Three-hop satellites are also of fairly frequent occurrence. On one occasion Baird (1954) recorded satellites on all echoes up to the five-hop, the separations decreasing progressively after the second.

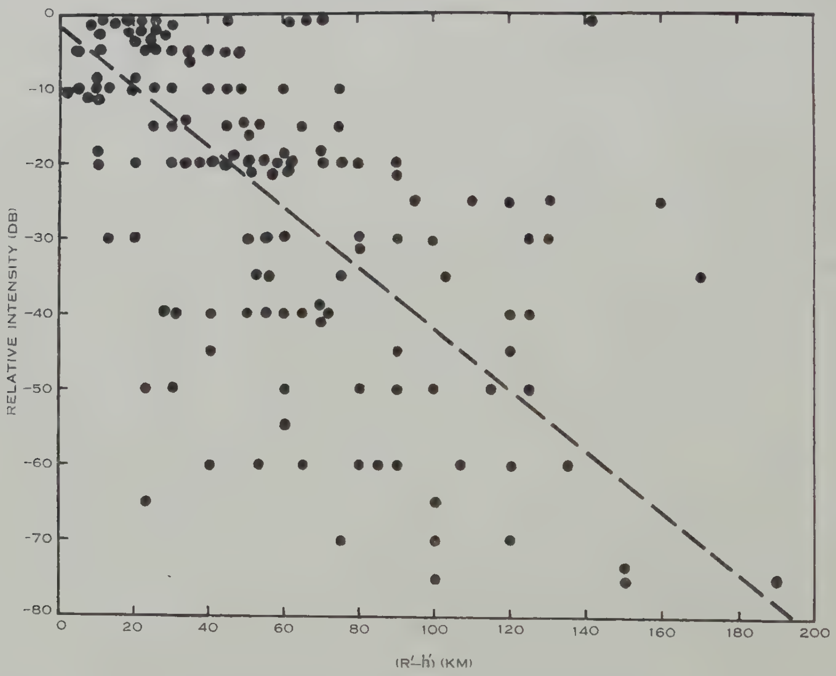


Fig. 14.—Relative intensity *v.* excess of range over minimum for satellites. Intensity taken as 0 dB at zero range excess. The broken line has a slope of $-0{\cdot}4$ dB/km.

(f) *Spatial Correlation of Satellites*

The $R't$ records at three stations 100 km apart are often closely similar (cf. Fig. 1). Not only does a multiplet at one station nearly always correspond to a multiplet at the others, but also 80–90 per cent. of individual satellites showing on one record (excluding those of less than 30 min duration) have counterparts on the other two. However, there are, at times, considerable differences in intensity, freedom from gaps, and duration of the individual satellites on the three records, especially when the multiplet contains many satellites. With stations located 300 km apart the correlation is noticeably lower, rather less than 30 per cent. of individual satellites showing definite correspondence.

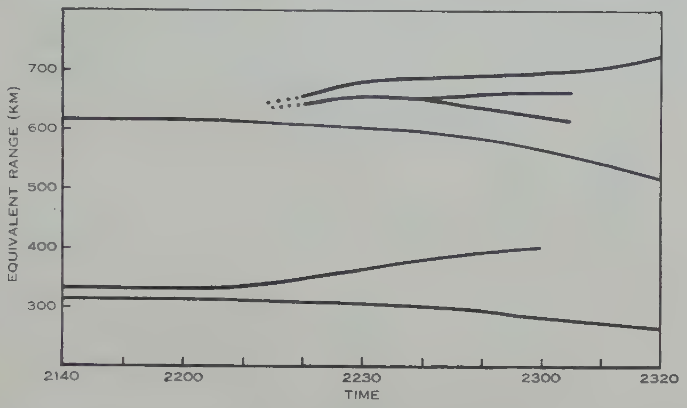


Fig. 15.—Re-plot of portion of *R't* records of Brisbane, Oct. 27, 1952, 2·28 Mc/s showing two (at times three) satellites on the double-hop trace.

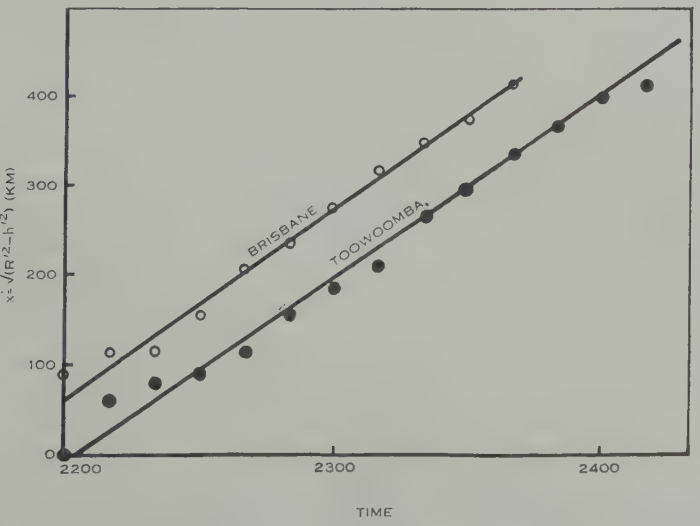


Fig. 16.—Plot of $x=\sqrt{(R'^2-h'^2)}$ versus *t* for corresponding satellites, Brisbane and Toowoomba, Sept. 4, 1952, at 2200 hr.

The *TR'* records showed markedly fewer and shorter satellites than the *TR* records at either end (cf. Plate 3, Figs. 1 and 2). Strohfeldt, McNicol, and Gipps (1952), recording at Brisbane pulses on 5·8 Mc/s from Camden 760 km away, found very few satellites.

Where correlation was established, it was found that the traces were usually displaced in respect to time, i.e. different values of t_0 were found. Figure 16 illustrates a case where such a displacement occurs. Figure 17 illustrates a case in which, in addition, dx/dt shows systematic variations, these occurring at different times at the two stations.

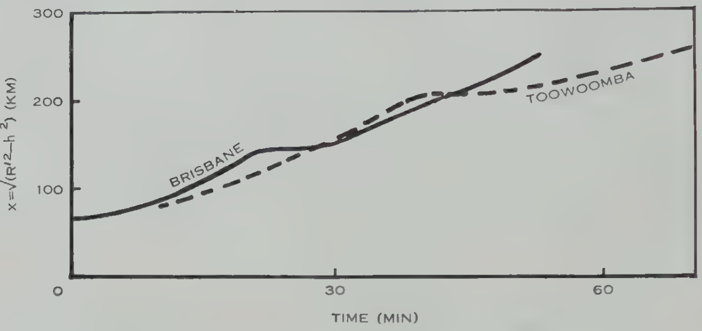


Fig. 17.—Plot of $x=\sqrt{(R'^2-h'^2)}$ versus t for corresponding satellites, Brisbane and Toowoomba, Aug. 21, 1952, at 2330 hr.

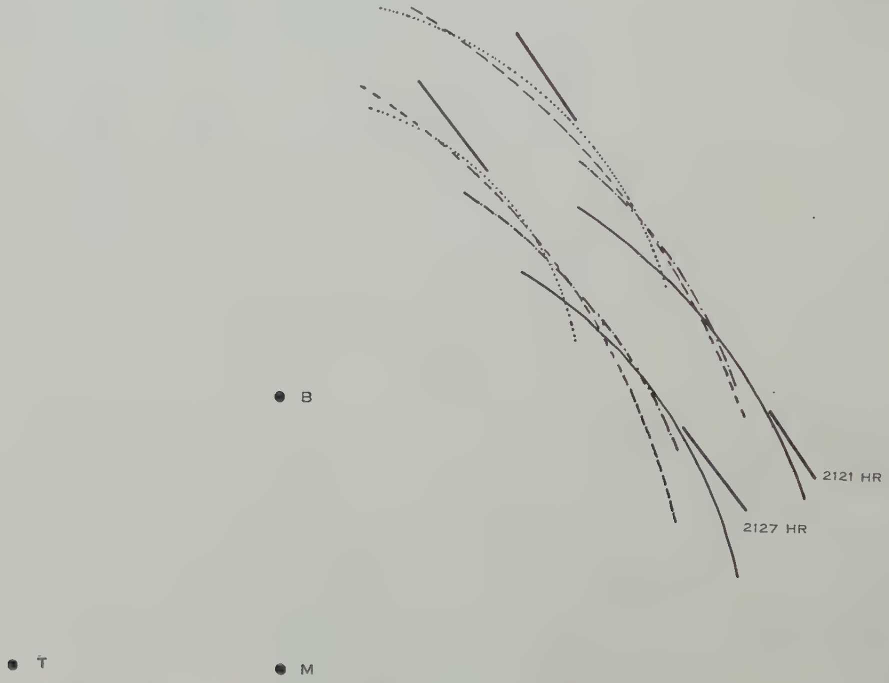


Fig. 18.—Plan of stations, with circles drawn to indicate simultaneous values of $x=\sqrt{(R'^2-h'^2)}$ for a satellite observed at three stations—Brisbane (M), Toowoomba (T), and Buderim (B), June 23, 1955. Full lines for Brisbane; broken lines for Toowoomba; dotted lines for Buderim. The " dot-dash " line is the ellipse deduced from TR' observations (2121 hr Buderim to Brisbane, 2127 hr Toowoomba to Brisbane). The straight lines indicate the common tangent in each case.

Where three records are available, it has been found that an important relationship exists between the values of x, namely, x_1, x_2, x_3, for the three stations at a given time t. This is most simply stated by reference to a diagram (cf. Fig. 18). If three points are marked in representing the plan positions of the three stations, and if circles are drawn round these points with radii corresponding to x_1, x_2, x_3, then the three circles have (fairly closely) a common tangent. There is rarely, if ever, any indication of a common intersection.

TABLE 2

CALCULATED AND OBSERVED AZIMUTHS FOR PARTICULAR SATELLITES

Time	Azimuth, φ	
	Calculated (deg)	Observed (deg)
2100	355	20
2227	0	327
2303	325	302
2315	329	320
2321	307	290
2345	355	339
2351	355	341
2342	313	307
2351	312	314
2354	197	200
2221	224	227
2315	307	311
0000	302	299
0154	326	328
0218	324	326
2230	37	25
0136	322	328
0200	322	326
0206	325	321
1954	289	283
2218	303	315
2118	316	329

This result (and also the x,t relationship) was discovered while investigating the validity of one of the hypotheses discussed in Part II of this series.

It is furthermore found that the normal to the tangent corresponds roughly in direction to the azimuth of arrival φ as determined directly. Table 2 gives a few typical results. A rough determination of velocity would thus be possible from observations from a single station (if azimuthal determinations are included).

In the simplest cases the common tangent moves roughly parallel to itself as time proceeds (cf. Fig. 18). In many cases, however, the direction changes somewhat; this would be true, for example, if the results given in Figure 17 were plotted in the way illustrated in Figure 18. Table 3 shows a comparison of the rate of motion of the common tangent with the corresponding values of dx/dt, as measured at Brisbane.

In the case pictured in Figure 18, a second satellite, not plotted, was also present. This gave a tangent moving in a different direction. The simultaneous occurrence in this way of two independently moving tangents is fairly common.

TR′ records were also available for the multiplet represented in Figure 18, and results calculated from them have been plotted in the form of ellipses on the figure. The diagram shows that the common tangents to the circles are also tangential to the ellipses:

TABLE 3

COMPARISON OF OBSERVED SPEEDS (AS DEDUCED FROM MOTION OF COMMON TANGENT) WITH VALUES OF dx/dt

Date and Time		Tangent Speed (m/sec)	dx/dt (m/sec)
15.vii.50	0130	90	78
22.vii.50	2130	107	103
23.vii.50	2300	48	51
23.viii.51	2000	48	56

$$\sqrt{(x_1^2+h'^2)}+\sqrt{(x_2^2+h'^2)}=2R'.$$

It is thus possible to plot the tangent without complete data from the three stations, provided *TR′* data are available from the station for which the *TR* data are missing.

The speeds and directions of movement of the common tangent for the cases fully examined* are shown in Figure 19. The most noteworthy feature is the tendency for the directions to group around 300°, and for the speeds to lie in the range 150–350 km/hr.

(*g*) *Duration of Multiplets and of Individual Satellites*

It is often impossible, especially when the multiplicity persists all night, to assign a definite beginning or end to individual multiplets, as the multiplets overlap. Broadly speaking, however, a typical duration for a mixed multiplet would be 1–2 hr.

It is somewhat easier to assign a value to the duration of individual satellites, although here also there are uncertainties (of the order of 5 min), because of the nature of the records. A histogram showing durations of such satellites for a series of periods distributed over June–November 1952, is shown in Figure 20. Satellites with durations less than 10 min have been excluded, for these are often obscured by the broadening of the trace. For the satellites shown, the mean duration is about 52 min. Figure 14 would suggest that an increase of, say, 10 dB in the overall sensitivity of the system would increase the measured durations, a rough estimate making the increase 10 min. However, the intensity usually falls off much more rapidly than 0·4 dB/km during the last few minutes of the life of the satellite and thus the actual dependence of measured satellite duration on equipment sensitivity is probably not very great.

---

* The authors are indebted to Messrs. Lamb and Woolcock for most of this analysis.

Corresponding satellite traces for different stations frequently show considerable differences in duration. Examination of a sample of 60 pairs of satellites which joined their respective main traces simultaneously (some junctions being found by extrapolation) showed about two-thirds with time differences of 10 min or more in their commencements (if convergent) or conclusions (if divergent). On the other hand, if pairs of convergent satellites starting simultaneously and pairs of divergent satellites finishing simultaneously were examined, differences of more than 10 min in the duration of the satellite were found in about one-half of the cases.

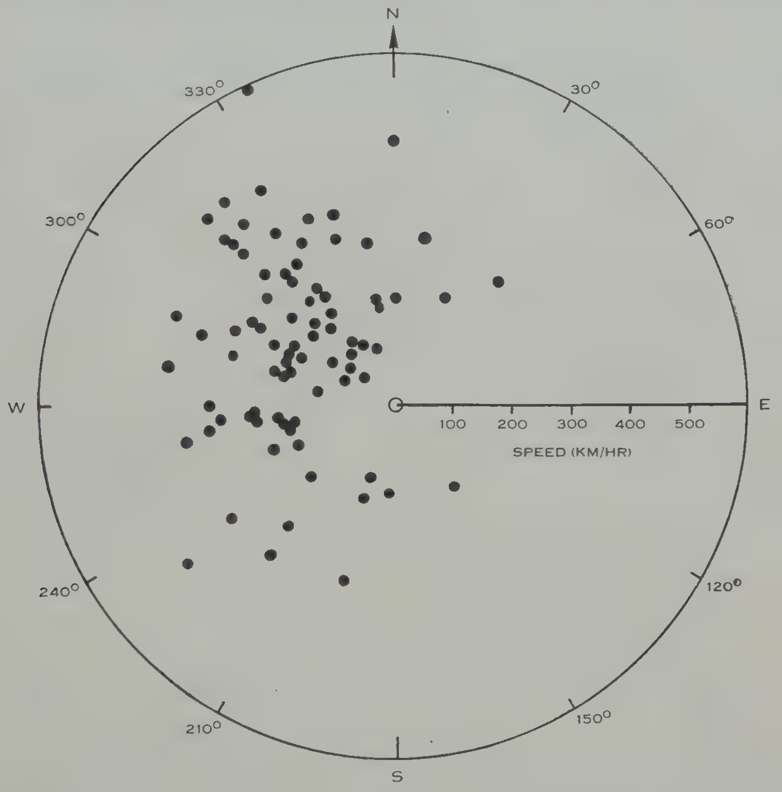


Fig. 19.—Polar plot indicating speed and directions of movement of the common tangent to the circles on the plan-position diagrams.

It was also noted that, considering pairs where the junctions were displaced by at least 10 min, about one-third were simultaneous, within 5 min, at the end of maximum range, while about one-quarter gave equal range separations, within 10 km, at that end.

A plot of durations against speeds (determined by dx/dt) for satellites is shown in Figure 21. There is evidently a tendency for the duration to decrease as the speed increases. It has been found that, as a rough approximation, the satellite terminates when x reaches a critical value, independently of the speed.

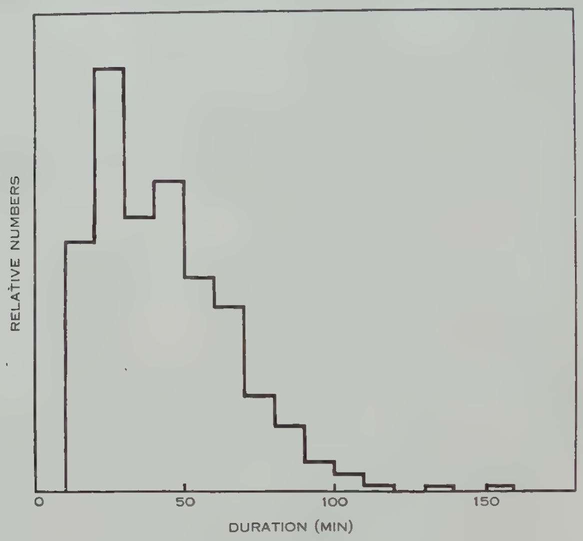


Fig. 20.—Histogram showing the distribution of durations for a large number of satellites in 1952.

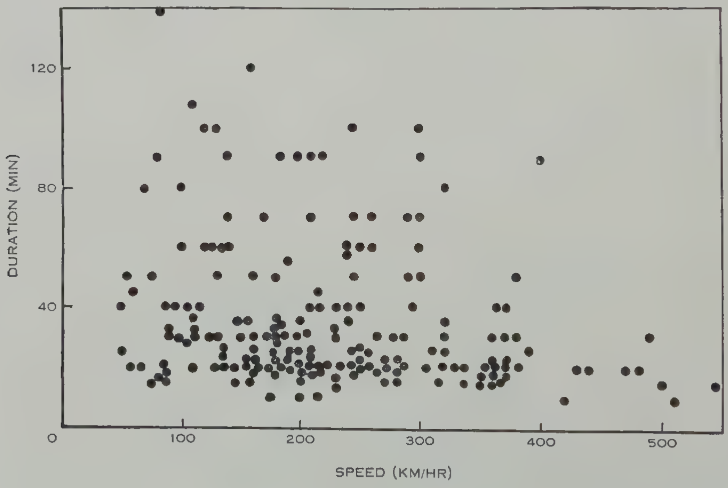


Fig. 21.—Plot of durations of individual satellites $v.$ corresponding values of dx/dt.

(*h*) Seasonal *Distributions of Range Multiplets*

The total duration per night (1900–0500 hr) of some degree of range multiplet structure, as observed on the fixed frequency records, has been examined for five nights in each month for a year, 1952–53. The following figures were obtained :

June–August (inclusive)	7 hr,
December–February (inclusive)	4 hr,
Other months in the year	2 hr.

The general agreement with the published "spread-F" curves of Gipps, Gipps, and Venton (1948) suggests that the spreading or diffuseness index has the same seasonal distribution as range multiplicity.

The number of occurrences of well-separated range doublets in each month for the years 1952–53* is plotted in Figure 22. The seasonal variation is clearly not large.

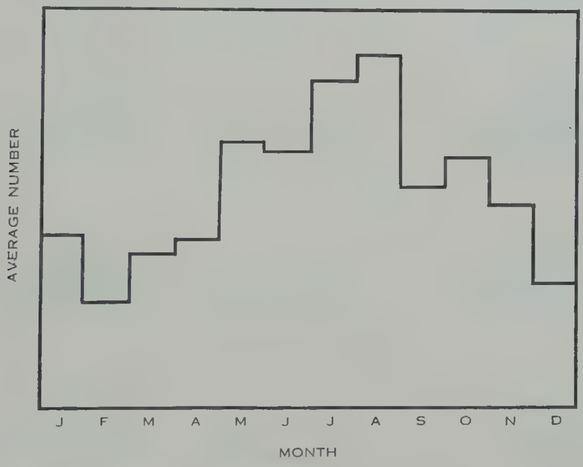


Fig. 22.—Seasonal distribution of occurrence of range doublets, 1952–53.

(i) *Diurnal Distributions of Range Doublets*

Since the number of range doublets in a single month is rather small, histograms have been constructed showing the total numbers of divergent, convergent, and mixed doublets, for each hourly period following actual ground sunset, during the six winter months, April–August (Fig. 23) and the six summer months, September–March (Fig. 24). Results for each year (1950–54 and 1952–53 respectively) are shown separately. The total numbers for the different years are not comparable, as for some years not all doublets were included, but only those of a more clear-cut character (usually greater duration).

The results suggest that, for winter months, the divergent doublets have a maximal probability of occurrence 2–3 hr after sunset, with a possible second maximum about 5 hr after sunset ; the convergent doublets have a maximum 3–4 hr after sunset, and the second (major) maximum an hour or two before sunrise ; and the mixed multiplets a single maximum around midnight. In summer, for which season information is available for only two years, for both divergent and convergent doublets the first maximum occurs at much the same

---

* The authors are indebted to Mr. D. Lamb for the 1952 analysis.

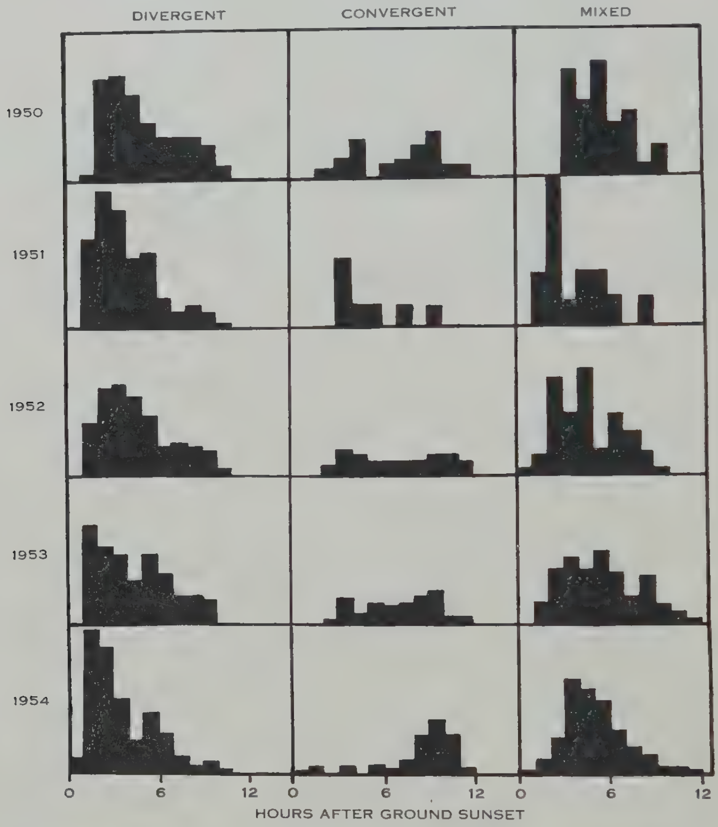


Fig. 23.—Diurnal distribution of range doublets, winter, 1950–54.

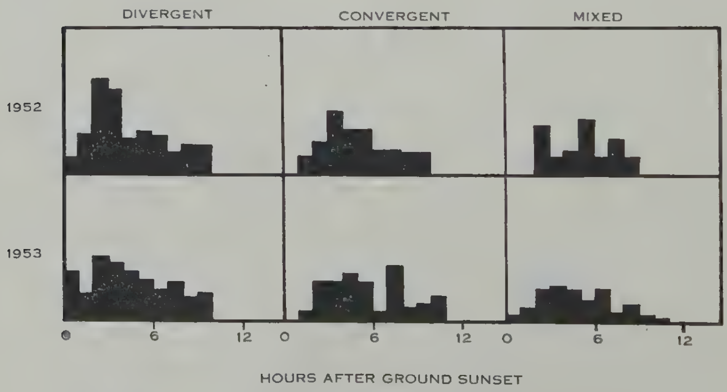


Fig. 24.—Diurnal distribution of range doublets, summer, 1952–53.

time as in winter, but the existence of a second maximum is doubtful. Data on mixed doublets in summer are insufficient to be significant.

(*j*) *Relationship of Satellites to Other* F-*region Behaviour*

The average equivalent height of the F region at Brisbane (as determined from the swept-frequency records) has been plotted in Figure 25 against hours after sunset. The averages were taken over six-month periods, for summer 1952–53 and winter 1953. On the curves are also marked the hours at which convergent and divergent satellites were most prevalent. It will be seen that divergent satellites occur most frequently on the rising portions of the curves, and convergent on the falling.

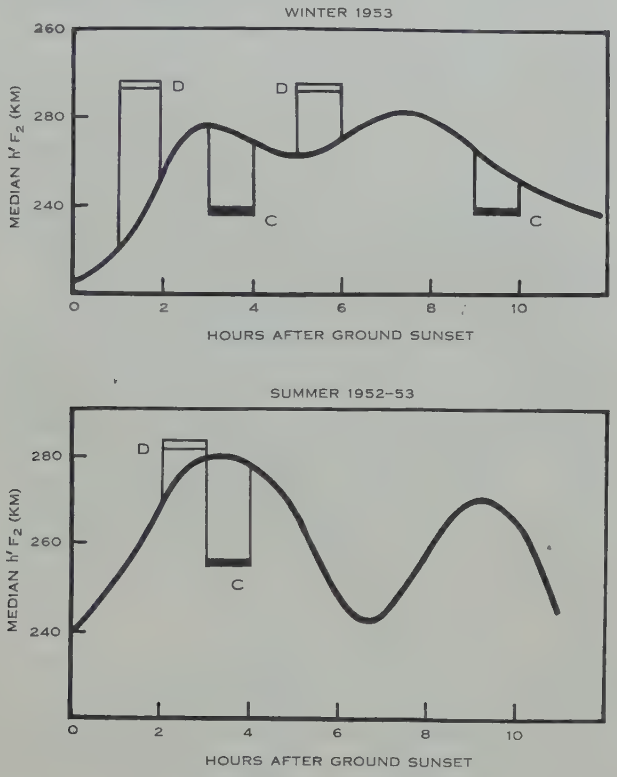


Fig. 25.—Diurnal variation of $h'F$ at Brisbane. Upper curve winter, 1953; lower curve, summer, 1952–53. Symbols attached indicate the positions of maximum frequency of occurrence of range doublets, open rectangles indicating divergent and filled-in rectangles convergent doublets.

It was found that selecting months over a period of 10 years when the 6-hr component of F height variation was prominent, the two maxima of frequency of occurrence of doublets of the two types showed out much more clearly.

The total electron constant of the F semi-layer has been determined* using the Kelso (1951) method for 10-min intervals during 12 range doublets, using the main swept-frequency trace. The changes were found to be small and not related in any systematic way to the behaviour of the doublet. Some investigations have also been made, with negative results, of possible correlations with changes of semi-thickness and of critical frequency.

---

* The authors are indebted to Mr. J. R. Hanscomb for carrying out this analysis.

(*k*) *Relationship between* F *Satellites and* E_S *Phenomena*

The equipments used provide data concerning the E_s layer similar to that provided for F. The fact that the seasonal variation of occurrence of F range multiplets is not dissimilar to the seasonal variation of occurrence of E_{sc} (McNicol and Gipps 1951) suggests that their diurnal variations might also be similar. No such correlation was found. However, if the diurnal variation of occurrence of three-hop E_s is compared with that of multiplets, some measure of inverse correlation is obtained. This is most clearly brought out by using the E_s results for July–August 1953, and the multiplet results for the same months (cf. Fig. 26). The results have been plotted as smooth curves, derived from the histograms by a " running means " method.

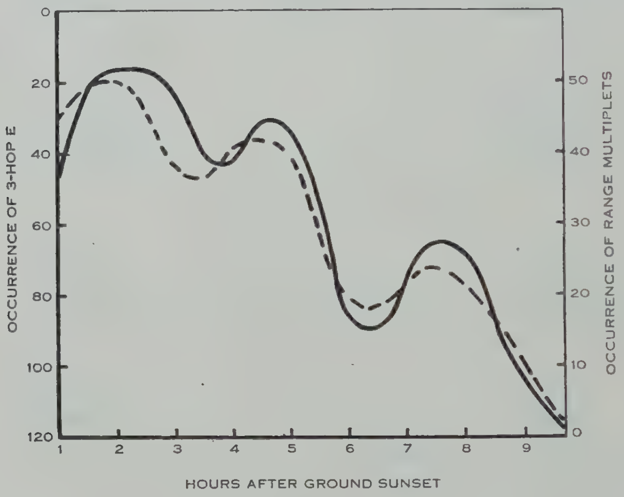
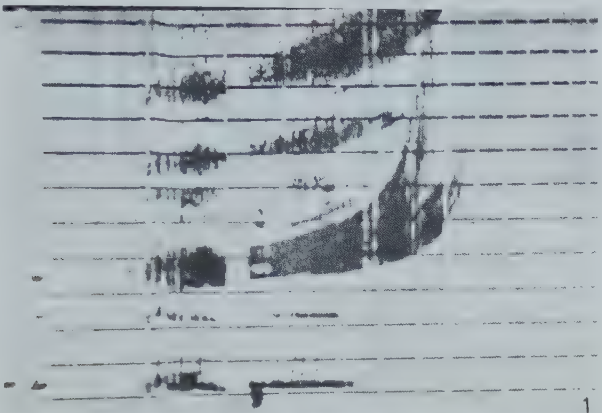


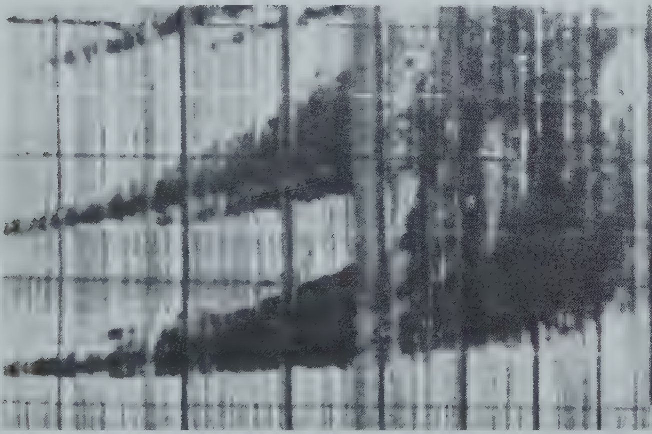
Fig. 26.—Inverse correlation between occurrence of range multiplets and $3E_s$ reflections. Diurnal variation of occurrence of three-hop E_s determined by number of 6-min periods for which it was present (full curve, plotted downwards). The broken curve gives frequency of occurrence of range multiplets (plotted upwards).

Echoes from the E_s region sometimes vary in range at an appreciable rate ; the traces usually show upwards curvature. If such curved E_s traces occur at a time when E_s traces of constant range are also present, the appearance of the E_s portion of the record is similar to that of the F region when it shows satellites. However, many isolated curved E_s traces also occur (cf. Plate 5, Fig. 1). In a few cases there is excellent simultaneity between F satellites and E_s curved traces, as in Plate 5, Figure 2, but the overall correlation, over a period of several months, of two years, was found to be less than 20 per cent. In fact, although some E_s was present during 65 per cent. of F satellites in 1952 (Lamb 1954),

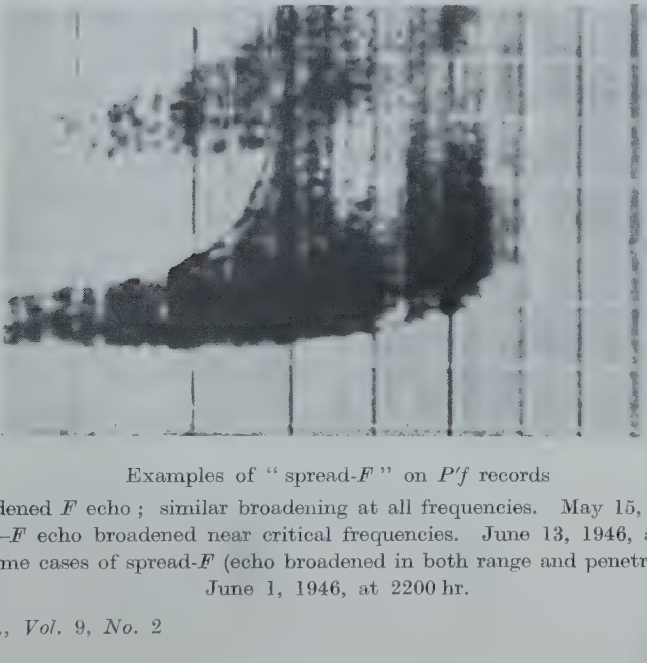
SPREAD-F IONOSPHERIC ECHOES AT NIGHT AT BRISBANE. I



1



2



3

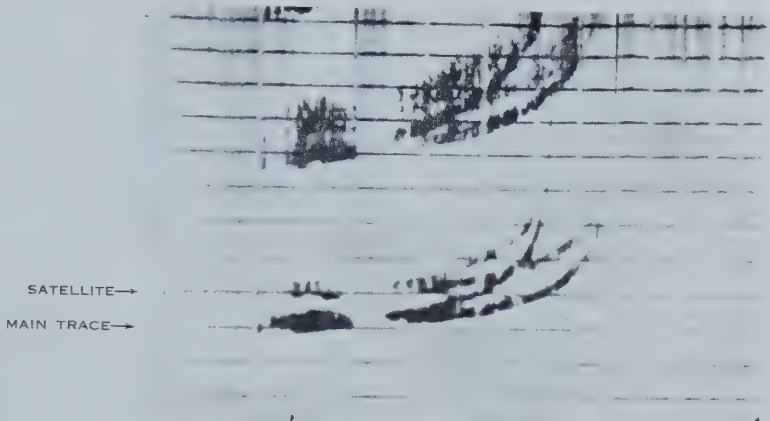
Examples of "spread-*F*" on *P′f* records

Fig. 1.—Broadened *F* echo; similar broadening at all frequencies. May 15, 1952, at 0215 hr.

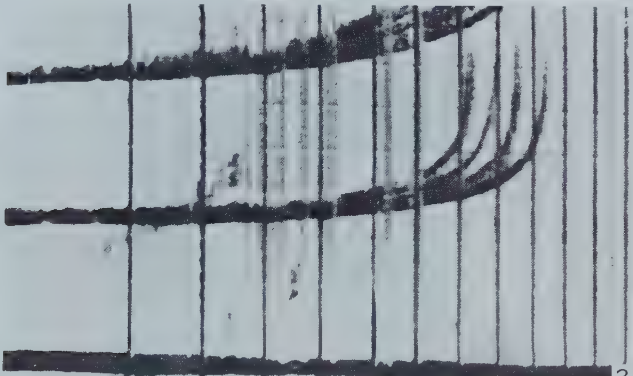
Fig. 2.—*F* echo broadened near critical frequencies. June 13, 1946, at 1840 hr.

Fig. 3.—Extreme cases of spread-*F* (echo broadened in both range and penetration frequency). June 1, 1946, at 2200 hr.

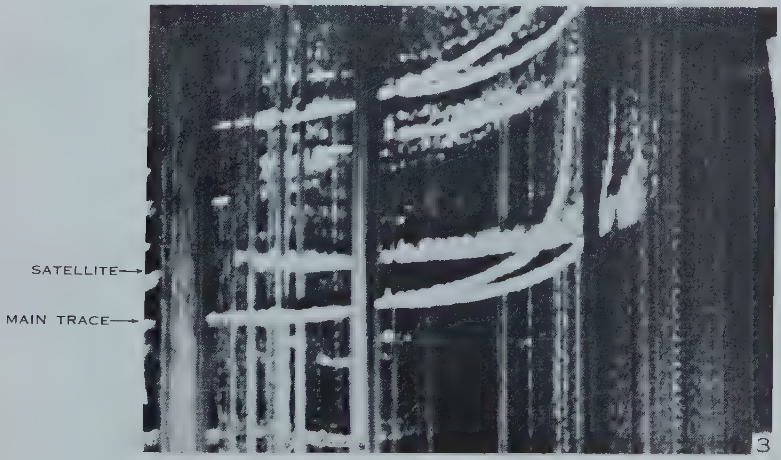
SPREAD-F IONOSPHERIC ECHOES AT NIGHT AT BRISBANE. I



1



2



3

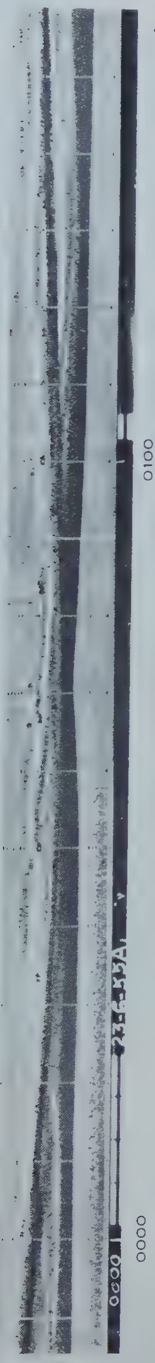
Examples of "doublets" on *P′f* records

Fig. 1.—Range doublet, both components having same critical frequency. October 11, 1952, at 0150 hr.

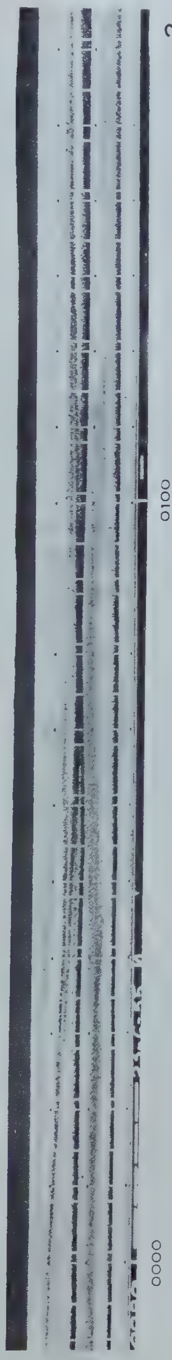
Fig. 2.—Range doublet, upper trace giving higher critical frequency. February 3, 1955, at 0210 hr.

Fig. 3.—Penetration-frequency doublet. December 17, 1955, at 2130 hr.

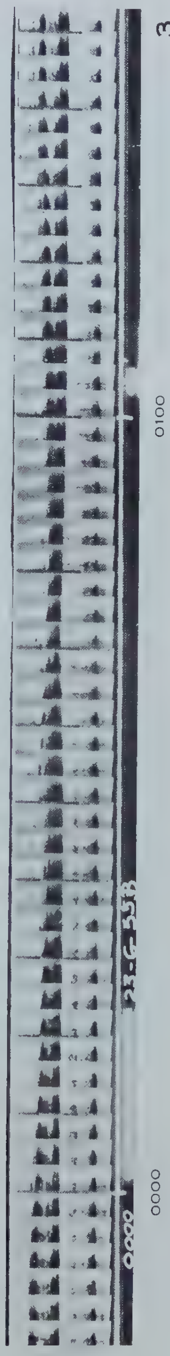
SPREAD-F IONOSPHERIC ECHOES AT NIGHT AT BRISBANE. I



1



2



3

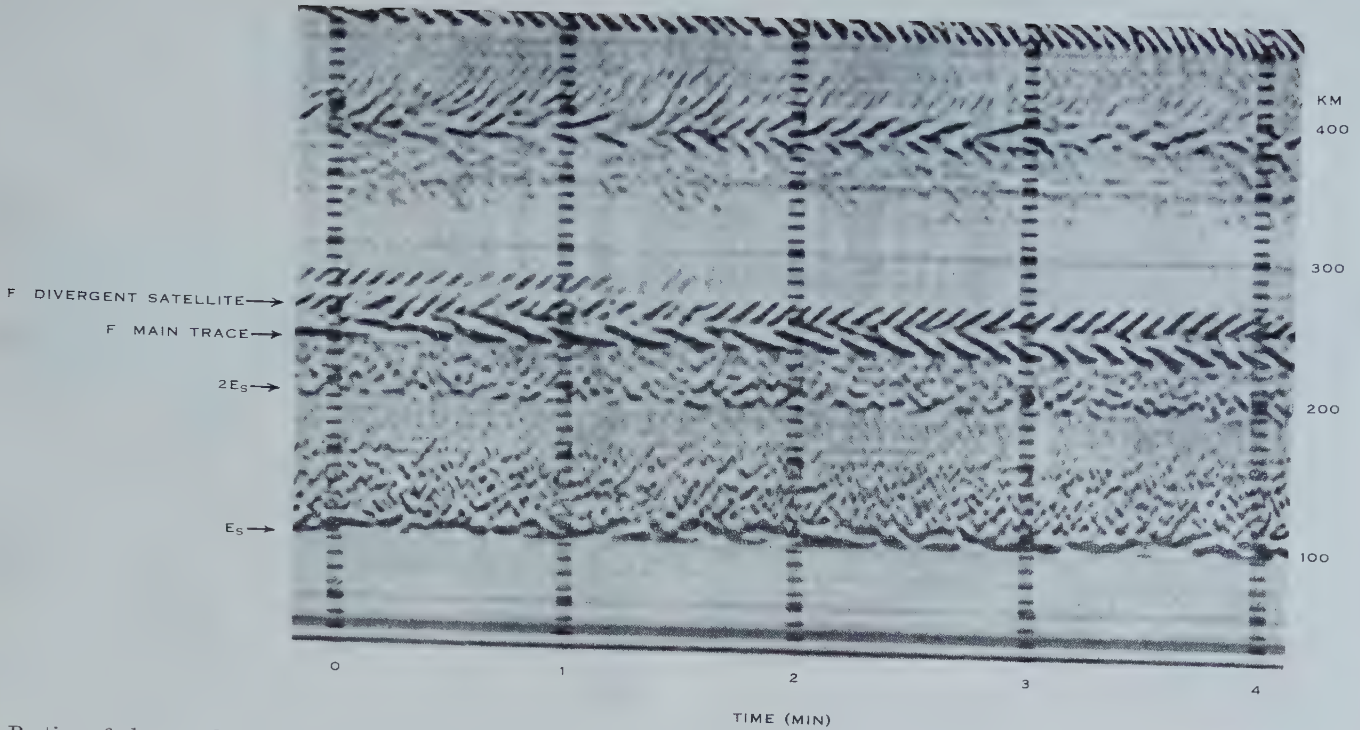
Mixed range multiplet (June 23 1955)

Fig. 1.—Record made at normal incidence at Brisbane, fixed gain.

Fig. 2.—Record made with pulses transmitted from Toowoomba, received at Brisbane, fixed gain.

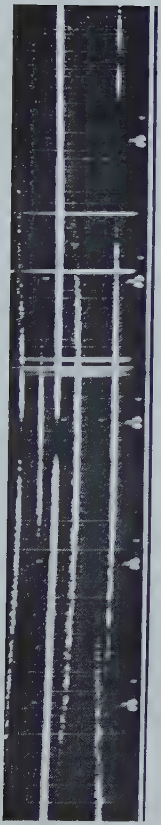
Fig. 3. Record made at normal incidence at Brisbane with swept-gain receiver.

SPREAD-F IONOSPHERIC ECHOES AT NIGHT AT BRISBANE. I

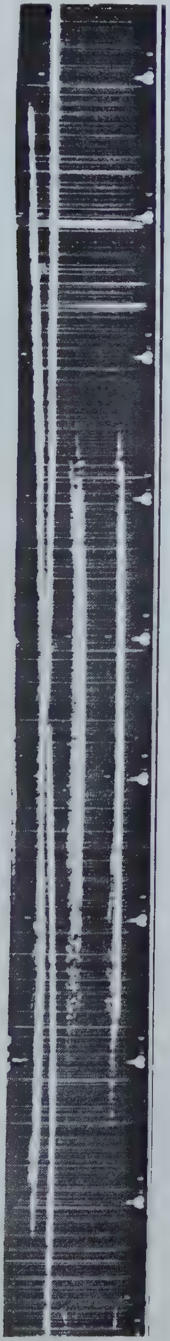


Portion of phase-path record, showing behaviour of main *F* trace and satellite during divergent range doublet. June 20, 1955, at 1900 hr. (The short lines sloping upwards to the right, which compose the satellite trace, indicate that the phase path is increasing; short lines sloping downwards to the right, which compose the main trace, indicate that the phase path of this trace is decreasing.)

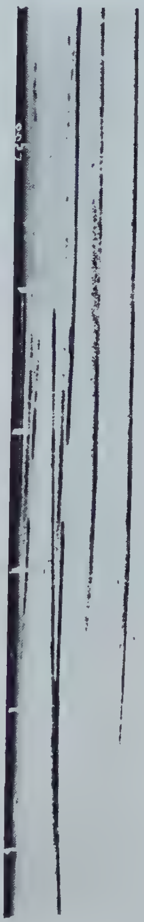
SPREAD-F IONOSPHERIC ECHOES AT NIGHT AT BRISBANE. I



1



2



3

Relations between F satellites and E_s phenomena

Fig. 1.—Strongly reflecting curved E_s traces, no F satellites. March 15, 1952, at 0100 hr.

Fig. 2.—Good simultaneity between curved E_s traces and F satellites. March 13, 1952, at 0230 hr.

Fig. 3.—E_s region blankets main F region trace, but not F satellite. April 2, 1952, at 2130 hr.

this is close to the ordinary probability of occurrence of E_s echoes on 2·28 Mc/s at night at Brisbane (Thomas, personal communication 1955).

On some occasions E_s region blankets the main F trace (2·28 Mc/s) but not the satellite (cf. Plate 5, Fig. 3), and vice versa.

(*l*) *Relationship to Geomagnetic Behaviour*

An investigation* of the relationship between occurrence of Brisbane range doublets and magnetic disturbance indices for Toolangi (lat. 37° 32′ S., long. 145° 28′ E.) led to negative results, in agreement with the findings of Uyeda and Ogata (1954).

IV. ACKNOWLEDGMENTS

This work forms part of the programme of the Radio Research Board and is published by permission of the Board. It has also been supported by the Research Committee of the University of Queensland. To these the authors wish to express their thanks. They wish also to express appreciation of the assistance rendered by students and technical staff, especially Mr. H. McGregor.

V. REFERENCES

BAIRD, K. (1954).—*Aust. J. Phys.* **7** : 165.

FINDLAY, J. W. (1951).—*J. Atmos. Terr. Phys.* **1** : 353.

GIPPS, G. DE V., GIPPS, D. I., and VENTON, H. R. (1948).—*J. Coun. Sci. Industr. Res. Aust.* **21** : 215.

JONES, R. E. (1953).—*Rev. Sci. Instrum.* **24** : 433.

KELSO, J. M. (1951).—Tech. Rep. No. 19, Ionosphere Res. Lab. Penn. State Coll.

LAMB, D. (1954).—M.Sc. Thesis, University of Queensland.

MCNICOL, R. W. E., and GIPPS, G. DE V. (1951).—*J. Geophys. Res.* **56** : 17.

MEEK, J. H. (1952).—*Nature* **169** : 327.

STROHFELDT, M., MCNICOL, R. W. E., and GIPPS, G. DE V. (1952).—*Aust. J. Sci. Res.* A **5** : 464.

THOMAS, J. A., and MCNICOL, R. W. E. (1955).—*Proc. Instn. Elect. Engrs.* **102** B : 793.

UYEDA, H., and OGATA, Y. (1954).—*Rep. Ionosphere Res. Japan* **8** : 103.

---

* The authors are indebted to Mr. W. M. Coleman for carrying out this analysis.

A STUDY OF "SPREAD-F" IONOSPHERIC ECHOES AT NIGHT AT BRISBANE

II. INTERPRETATION OF RANGE SPREADING

By R. W. E. McNicol* and H. C. Webster*

[*Manuscript received January* 31, 1956]

Summary

A review is given of various hypotheses which might explain the behaviour of "range multiplets", in which satellite echoes accompany the F_2 echo at frequencies well below the critical penetration frequency. It is shown that these cannot arise from stratification of the layer. Some few may be due to reflection from the F layer of waves previously scattered upwards by an E_s irregularity. The great majority, however, appear to originate in reflections from irregularities in the F layer itself. This is confirmed by the directional observations, multiple-hop echo records, correlations between separated stations, and the usual shape of the range-time curves, as described in Part I.

Evidence is presented indicating irregularities with fronts up to several hundred kilometres long, with curvatures ranging up to 10^{-2} km^{-1}. The irregularities appear to take various forms, of which the simplest are (*a*) an inverted trough, around 100–200 km wide, (*b*) a gap, of roughly the same width, with reflecting surfaces which have, at some points, curvatures up to at least 10 km^{-1}, (*c*) a step, with height and breadth not exceeding some 10–20 km, again with edges which are strongly curved in parts. Each irregularity probably persists for a period of the order of 1–2 hr and is usually propagated through the F region with a nearly uniform velocity of the order of 200 km/hr. The balance of the evidence indicates that this is not related to any general drift of the region.

I. INTRODUCTION

The range multiplets described in the preceding paper (herein referred to as Part I) would appear, *a priori*, to admit of three possible interpretations :

(1) A real stratification of the F layer leading to reflections at different heights.

(2) Double reflections, one reflection (or scattering) at a non-horizontal ionization contour of the E_s layer, and one at F.

(3) Reflections from non-horizontal ionization contours in the F layer, displaced laterally from the zenith point over the station, together with vertical reflection from horizontal contours directly overhead.

In this paper these three hypotheses are examined in the light of the results assembled in Part I and it is shown that, in the great majority of cases, hypothesis (3) appears the most plausible ; this is in accord with the conclusions of other

---

* Physics Department, University of Queensland.

recent workers. Hypothesis (2) may operate in a very few cases. Using hypothesis (3) certain conclusions can be drawn concerning the horizontal irregularities in the ionization contours.

II. The Stratification Hypothesis

On this hypothesis, the satellite echoes are returned from layers above the normal F_2; the implication is that they come down vertically, or nearly so. The directional observations, which indicate that, for satellites of large range separation, the signals arrive at a considerable angle to the vertical, thus rule out this hypothesis for these satellites. For satellites close to the main trace the direction cannot be determined, but it seems improbable that the hypothesis is valid under any conditions for the following reason. Irrespective of the separation, the main trace on the swept-frequency ($P'f$) records has the same appearance whether satellites are present or not. This trace suggests a thick layer of nearly uniform consistency (for variations in maximum ion density would lead to spread in critical frequency, which is not characteristic of range multiplets). Hence it is difficult to see how any of the wave can penetrate through the layer. Considerable penetration would be necessary to enable an upper layer to give an echo of comparable intensity, as is observed. Moreover, considerable group retardation, increasing markedly with frequency, would occur in passing through the lower layer; one would thus expect the separation of the traces to increase as the frequency rises; this does not occur. There are thus essential discrepancies.

III. Hypotheses Involving Sporadic E

Eckersley (1937) has suggested that spread-F could be due to scatter from irregularities in the E_s layer, followed by regular (oblique) reflection at the F layer. The type of path postulated is shown in Figure 1 (*a*). Beckman, Menzel, and Vilbig (1941) have obtained results consistent with this hypothesis; but Dieminger (1951) concludes that it is untenable. On the other hand Meek (1952) has obtained clear evidence of the occurrence of the phenomenon at high latitudes. Several of the general features of range multiplets, as set out in Part I, are hard to reconcile with this hypothesis, namely:

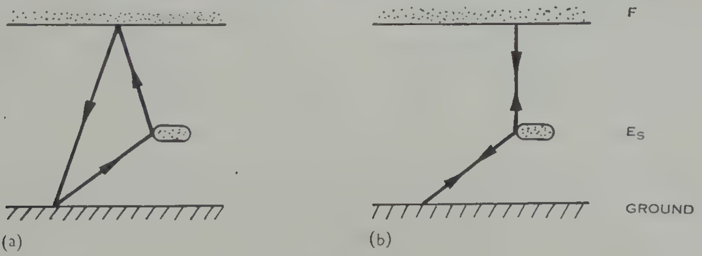


Fig. 1.—Possible F satellite ray paths via E_s clouds.

(*a*) The equality of critical frequencies of satellite and main trace. Since the F reflection in Figure 1 (*a*) is oblique, the critical frequency should be raised in the ratio sec $\psi : 1$, where ψ is the angle of incidence on the F layer. From the heights of E_s and F the dependence of ψ on R' can be found ; the increase in critical frequency should be evident as soon as $R'-h'$ exceeds 50 km. There are, in fact, a few cases where the critical frequency of the satellite is markedly higher than for the main trace ; these will be dealt with in Section VI.

(*b*) The almost universal agreement (shown in Table 1 of Part I) between the measured value of downcoming angle θ_M, and its value θ_C calculated from

$$\cos \theta_C = h'/R',$$

where R' is the actual, and h' the minimal range. In Figure 1 (*a*) the values of θ_1 and θ_2 would be considerably different from θ_C, and, although it is not unexpected that θ_M should be intermediate between θ_1 and θ_2 (since the path shown in Figure 1 (*a*) could be traversed in either direction), the agreement is too close for this explanation to be plausible. In a very few cases θ_M is in reasonable agreement with θ_1 or θ_2.

(*c*) The systematic changes in the value of R' for the main trace near the junction with a satellite (Fig. 9 of Part I). Gipps (1954) suggested that for mixed multiplets these changes could be due to group retardation in E_s. The normality of the shape of the $P'f$ curves, as contrasted to the distortion usually associated with group retardation by a lower layer, makes this explanation improbable and the absence of any pronounced difference between group path and phase path (where recorded) appears to rule it out.

(*d*) The high intensity frequently encountered in satellites, especially for range separation less than 40 km (cf. Fig. 14 of Part I), at times when the $(F+E_s)$ echo is absent or quite weak. If the satellite, like the $(F+E_s)$ echo, involves one reflection each from F and E_s, it should be roughly comparable in intensity.

(*e*) The absence of clear correlation between F satellites and the curved E_s traces referred to in Section III (*k*) of Part I. It is reasonable to suppose that, if there is sufficient forward scatter from E_s irregularities to give an overall reflection coefficient for the F satellite within an order of magnitude of that for the main trace, the back scatter will be not many orders lower.

Since there is some 30–40 dB of sensitivity in hand, a curved E_s trace should be recorded. Although occasionally there is very good correlation between F and curved E_s traces it is normally not much better than would be expected from pure chance. The mechanism for the inverse correlation (cf. Fig. 26 of Part I) between diurnal variation of three-hop E_s and that of occurrence of range multiplets is too obscure for it to provide evidence either way.

(*f*) The absence of correlation with geomagnetic disturbances. The currents responsible for these disturbances flow close to the E_s layer and might be expected to lead to ionospheric disturbances and hence to scatter clouds.

The hypothesis can only be valid for those cases where the critical frequency of the satellite is appropriately greater than that of the main trace. For short-lived satellites, of small separation, the change in critical frequency may be undetectable ; hence such satellites may possibly have such an origin. If so, it would be expected that the diurnal variation of occurrence of short-lived satellites would be less marked than that of long-lived, for there is little consistent diurnal variation of occurrence of E_s clouds (Thomas, unpublished data). The experimental data are not sufficiently extensive to test this point. Since the E_s scatter clouds may be quite small, this explanation would be consistent with the poorer correlation between short-lived than between long-lived satellites recorded at different stations.

With a few longer-duration satellites, records of the type shown in Plate 2, Figure 3, of Part I, have been obtained. The process shown in Figure 1 (*a*) should give a definite relationship between critical frequency rise and range. The experimental results are not in very good agreement, in the few cases which have been checked. Further, for oblique incidence, the $P'f$ curve for the satellite should have a characteristic " bent-back " configuration ; this is not shown in the case under discussion and is rarely shown in other similar cases. Since the third hypothesis (F irregularity) can readily explain the critical-frequency differences, these discrepancies are serious.

Angle-of-arrival measurements and the shape of the traces also suggest that E_s scatter gives a few F satellites each year at Brisbane. The total evidence is, however, far from conclusive, and the main bulk of satellites are certainly not due to this process.

A variation of the hypothesis is indicated in Figure 1 (*b*). It would appear to be less probable than the first and no definite evidence has been established of its occurrence at Brisbane. In this case the critical frequency should be normal, but the zenith angle should be larger than $\sec^{-1}(R'/h')$ and a curved E_s trace should exactly parallel the F satellite.

IV. The F-irregularity Hypothesis

Gipps, Gipps, and Venton (1948) from the earlier observations at Brisbane concluded that the satellites arise from reflections from portion of the ionosphere at which the gradient of ionization is appreciably inclined to the vertical ; these regions are displaced horizontally to an appropriate extent from the zenith but are at the normal height. Later work, in particular that of Bramley and Ross (1953), Bibl, Harnischmacher, and Rawer (1955), and Ratcliffe (1955), has shown that such non-vertical gradients are of common occurrence. This hypothesis is clearly consistent with the broad general conclusions of Part I, in particular :

(*a*) The general relation between the downcoming angle, θ, and R', namely,

$$R'=h' \sec \theta.$$

This follows by simple geometry from the hypothesis.

(*b*) The quantity x defined by

$$x=\sqrt{(R'^2-h'^2)},$$

varies roughly uniformly with time. This follows if one assumes, as seems reasonable, that the irregularity moves horizontally with a sensibly constant velocity.

(*c*) The existence of the common tangent in the plot using observations at three stations and the parallel motion of the tangent. If one extends the present hypothesis, by assuming that the irregularity is a ripple or front having considerable extent and constant configuration perpendicular to the direction of motion, then the common tangent locates this front and the experimental behaviour follows automatically.

(*d*) The agreement between the measured azimuth and that deduced from the position of the common tangent. This also follows by simple geometry from the extended hypothesis.

(*e*) The appearance of double-hop satellites, the similarity of separation of the double-hop and single-hop traces, and the occasional appearance of a second satellite. This is readily understood from Figures 2 (*a*) and 2 (*b*) which show how the two satellites can be formed.

(*f*) The appearance of three-hop and higher-multiple hop satellites and their separations. The scheme for three-hop reflections is indicated in Figures 2 (*c*) and 2 (*d*); higher multiples follow corresponding patterns. The separations (in the few records measured) are fully consistent with this scheme.

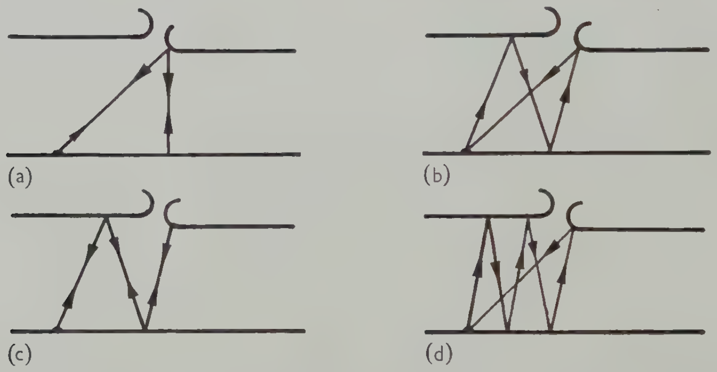


Fig. 2.—Possible paths for two-hop and three-hop *F* region satellites.

V. DETAILED CONFIGURATION OF THE IRREGULARITY

If the suggested hypothesis is the correct one, it should be possible, without passing the bounds of plausibility, not only to account for the idealized behaviour summarized in Part I, but also to explain departures from this behaviour. To do this, it is clearly necessary to make specific assumptions concerning the ionization contours.

(*a*) *Inverted Trough*

In his explanation of day-time *F* satellites, Munro (1953) postulated an inverted trough configuration. Following this idea, let us consider an ionospheric irregularity defined by

$$z=a+b(1+\cos 2\pi x/\lambda),\ \text{when}\ -\tfrac{1}{2}\lambda<x<+\tfrac{1}{2}\lambda,$$
$$z=a\ \text{otherwise.}$$

Here z represents the height of a particular ionization density;

x represents horizontal distance along the direction of movement (i.e. perpendicular to the "front"), measured from the (moving) centre of the irregularity;

a, b, λ are arbitrary constants, a only being dependent upon the particular value of ionization density assumed. (In this respect the picture is simpler than that used by Munro.)

It is assumed also that z is independent of the third coordinate y (perpendicular to x, z). This is also implicit in the hypothesis of Toman (1955).

It is reasonable, on this assumption, to treat the problem by geometrical optics, regarding the particular ionization contour appropriate to the frequency in use, as a simple reflecting surface. On this basis and taking $a=230$ km, $b=20$ km, $\lambda=100$ km, and the speed as 450 km/hr, an $R't$ pattern of the form shown in Figure 3 is to be expected. By comparing it with Figure 8 of Part I, there is seen to be a considerable measure of agreement.

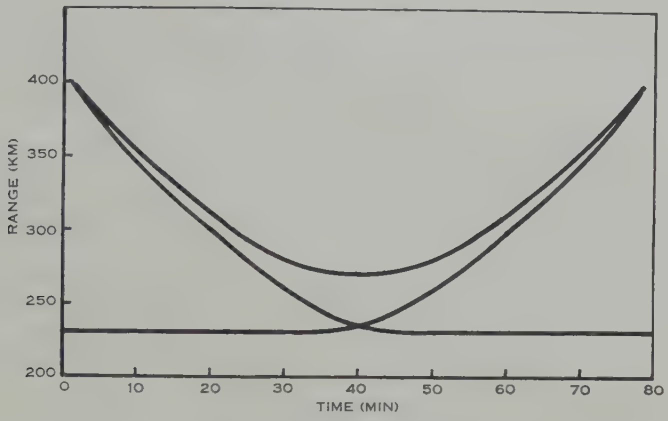


Fig. 3.—Expected $R't$ pattern for " inverted trough " disturbance travelling horizontally in F region.

The chief discrepancies are : (i) the rise of the lowest trace, to the point of intersection is larger experimentally than Figure 3 predicts ; (ii) the pattern is unsymmetrical, the height being lower at the end than at the beginning; (iii) the upper member of the pattern is absent for much of the period. The first of these discrepancies can be adjusted by assuming a more abrupt rounding-off at the edges of the trough than has been specified. The second can similarly be adjusted by assuming an asymmetrical, instead of symmetrical, relationship between z and x, while the third is discussed in the next section.

(b) *The Gap*

The absence of the upper member of the pattern leads to the conclusion that part of the upper section of the trough is non-reflecting, possibly due to strong absorption immediately below the reflection level. Other records of mixed doublets exhibit this discrepancy even more markedly. (Figure 8 of Part I is a rare case, selected as showing the upper member clearly.) It is therefore appropriate to conceive the geometrical reflector as having a gap, in some cases with a few " islands " within the gap. The rounded-off edges of the gap supply the principal intersecting traces, and the islands supply the vestiges of the top member, and perhaps the (poorly correlated) subsidiary satellites which frequently appear in a mixed multiplet. Figure 4 illustrates the pattern of satellites expected when there is a single island, at normal height, in a gap 200 km wide. (This is the order of magnitude found experimentally.)

Complex patterns bearing a resemblance to this are often observed. In such cases, the resolution is usually insufficient for the direction-of-arrival equipment to function.

It may be pointed out that the linear relationship between x and t suggests that the edges of the gap have a large curvature. The experimental points do not, however, fit such a relationship sufficiently closely to justify any firm conclusion. This matter is taken up again in Section VII.

In the production of a second satellite to the two-hop F trace by the mechanism pictured in Figure 2 (*b*), reflection from the (level) layer to one side of the irregularity is required. No reflection is possible from within the " gap " postulated for a mixed multiplet and this satellite should therefore be absent for some time before and after the intersection of the two principal traces. In the one case suitable for detailed examination the satellite appeared first some time after intersection.

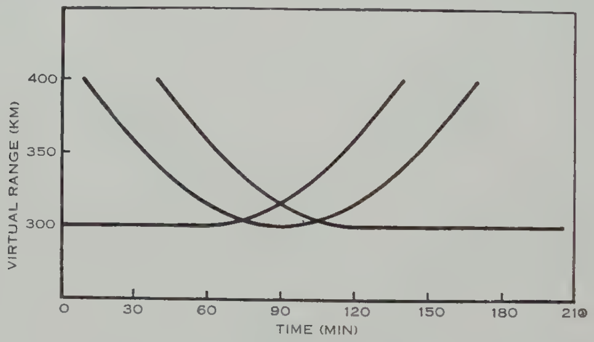


Fig. 4.—Expected $R't$ pattern for disturbance consisting of gap 200 km wide with central cylindrical cloud, travelling horizontally in F region.

The geometrical picture can obviously only serve as a general guide. With curved ionization contours, some group-retardation effects are probable; these are presumably responsible for the discrepancies between phase- and group-path changes found to occur when there is rapid change.

(*c*) *The Step*

As with the inverted trough, it is reasonable to suppose that sometimes the relevant ionization contour is at different heights on the two sides of the gap, as suggested in Figure 5. In fact, if the gap arises from some sort of cellular wave (Martyn 1950) there would be a tendency to upward drift on one side and downward on the other. Under the conditions pictured in Figure 5, where the height difference is comparable with the gap width, the range doublet would clearly be markedly asymmetrical. The region immediately below the reflecting contour is highly absorbing and would tend to prevent the appearance of the satellite even to the right of the point X. In many mixed doublets, the satellite on one side of the junction is, in fact, very short.

The existence of pure convergent and divergent multiplets can be fairly satisfactorily accounted for by postulating either such a narrow gap, or a " step " in the reflecting surface, as indicated in Figure 6 (*a*). This may be regarded as the limiting case of gap, when the width becomes negligibly small. Figure 6 (*b*) indicates the anticipated $R't$ record when the step is moving in the direction indicated with a speed of 200 km/hr ; reversal of velocity changes the convergent into a divergent satellite. The theoretical curve indicates a short trace (shown dotted) on the far side of the intersection. This is absent in the actual records ; this could be due to confusion with the relatively broad lower trace, or to refraction associated with the strongly curved ionization contours below the reflection surface. It will be seen from the diagram that this picture also accounts for the records, noted in Part I, in which the single trace on one side of the junction is continuous with the satellite trace rather than the main trace

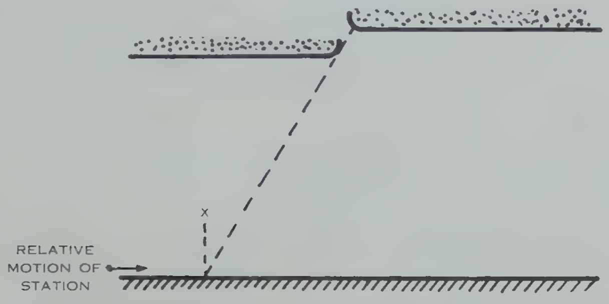


Fig. 5.—Narrow gap in F region—heights on two sides of gap unequal. No satellite visible if recording station is to left of X.

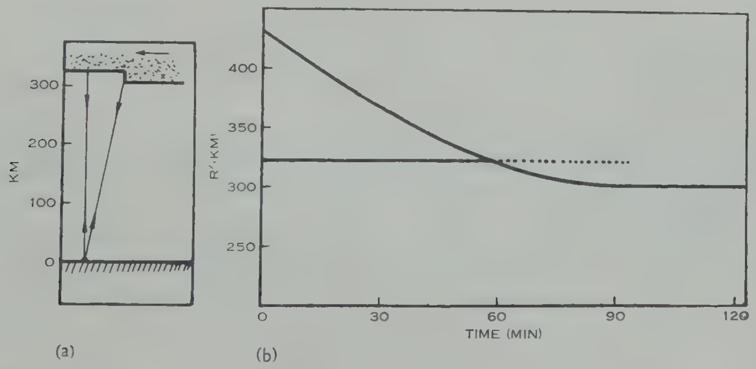


Fig. 6 (*a*).—Step F region disturbance.
Fig. 6 (*b*).—Expected $R't$ pattern for a disturbance in the form of a step 20 km high with speed 200 km/hr.

on the other side. It also explains the tendency for the "main" trace to rise at the junction with a divergent, and fall at the junction with a convergent satellite (Fig. 10 of Part I).

The satellites that appear at lenticels can probably be fitted into a scheme of this sort. Detailed checking is difficult because it is clear from the splitting of o- and x-rays and the wide discrepancy between group-path and phase-path changes that, in this case, successive ionization contours are by no means parallel or similar and that the simple geometrical picture is not valid.

(*d*) *Horizontal Extent and Horizontal Curvature of the Edges*

A rough estimate of the length of the gap or step can be formed from the falling-off of individual satellite correlation as between Brisbane-Toowoomba (95 km with 80–95 per cent. correlation) and Brisbane-Goondiwindi (300 km with 20–30 per cent. correlation). The orientation of the edge, in some of the cases giving correlation between Brisbane and Goondiwindi, was roughly along the line joining these stations; some steps, at least, must therefore have been more than 300 km long. Where the orientation was more nearly at right angles to this line, the correlation is more a measure of the permanence of the irregularity; this aspect will be discussed later. As it has not been possible to determine orientations in those cases where correlation was absent, the information is inconclusive, but it suggests that the average length of the edge is comparable with 300 km.

If the edges of the gap, or the step, as the case may be, are straight along their whole length, it is clear that, provided they move with uniform speed perpendicular to their length,

(i) the plot of x versus t will give a straight line; and

(ii) the slope of this line will be the same for all the observing stations.

Although certain satellites satisfy these requirements fairly closely, some (cf. Fig. 17 of Part I) show quite systematic departures. Such departures can be simply accounted for by postulating that the edge is curved (in a horizontal plane). If the edge retains its shape as it moves, this implies that any particular part of the edge will not, in general, move along its normal (cf. Fig. 7). Since the speed v' measured by dx/dt is the normal component of velocity, it follows that

$$v' = v \cos \alpha,$$

where v is the true velocity and α the angle it makes with the normal. As a curved edge moves, the reflecting point, for a particular station, will move along the edge and the value of α will change. Moreover, α will have different values for different stations.

Under these conditions the normals at two or more points on the edge may pass through a given station, so that one edge could provide several satellites, with different values of v'. It is, however, more likely that, when several satellites with different slopes coexist, they arise from a pattern of steps or gaps with different orientations and different (normal) velocities.

With these supplementary postulates, allowing steps, gaps, with and without intermediate clouds, and permitting the edges of steps, gaps, and clouds to be curved horizontally, sufficient arbitrary features are available to explain the tracks in all multiplets which have been recorded.

(e) *Implications of the Intensity Measurements*

Although, as discussed in Part I, the changes of intensity during the course of a satellite follow no regular pattern, the average behaviour is moderately well represented by a decrease at the rate of 0·4 dB/km, as indicated by the line marked E in Figure 8.

This is difficult to reconcile with the simple geometrical model, for the intensity of reflection from a cylindrical (convex) mirror of radius between 1 and 20 km, calculated on a geometrical basis, should decrease with increasing distance roughly as shown in curve A. (In computing this curve it was assumed that the "aperture" of the mirror is small compared with the distances of source and image.) The experimental points shown in Figure 14 of Part I nearly all fall well below this curve.

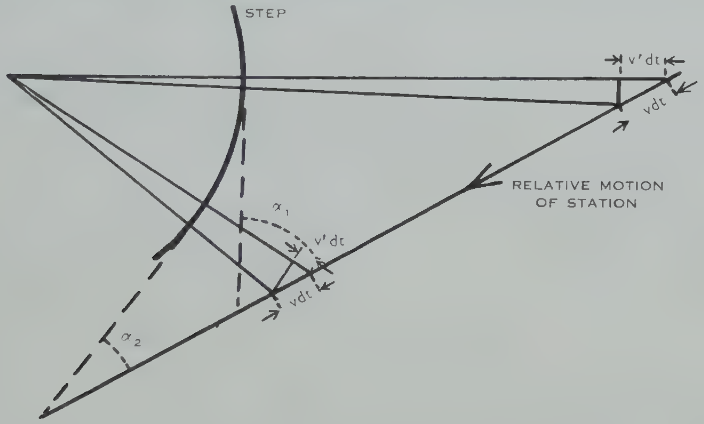


Fig. 7.—Dependence of dx/dt on orientation of tangent for step curved in horizontal plane.

In seeking an explanation of the discrepancy, calculations were made assuming the radius of curvature of the edge to be small compared with the wavelength (132 m). Diffraction theory is then necessary and as a first approximation Fresnel's theory for diffraction at the edge of an obstacle is applicable (regarding the ionosphere on one side of the edge as a semi-infinite reflector). Curve B was obtained thus; it gives a little better fit but falls too rapidly initially and too slowly later.

Since the ionosphere is known, from fading studies etc. to be far from smooth (cf. Briggs and Phillips 1950), it seems more reasonable to calculate intensities upon the same basis as is used by Briggs and Phillips in their analysis: the power scattered back from the portion of the ionosphere subtending a small solid angle dΩ at the source, at an angle θ with the vertical is assumed given by

$$dW \propto \cos^n \theta d\Omega,$$

and the contributions of the various elements are assumed additive arithmetically. For the ionosphere on one side only of the edge (assumed to have a radius of curvature small compared with the wavelength) the intensity variation has been calculated on this basis, taking $n=200$, which is typical for the F region. The result is indicated in curve *C*, of Figure 8.

The experimental points of Figure 14 of Part I nearly all fall between curves *A* and *C*, filling this area nearly uniformly. A plausible explanation of this result is that the curvature of the edge is not uniform, but changes considerably within one section, and from one section to another, along the length, so that there occur radii of curvature both large and small compared with the wavelength. This is reasonable for an ionosphere possessing the amount of roughness indicated by typical fading records. The maxima and minima, shown in Figure 13 of Part I, could be due to movement of the reflecting point along the edge, between regions of different curvature; it could also possibly arise from interference of reflections at two or more points along the edge.

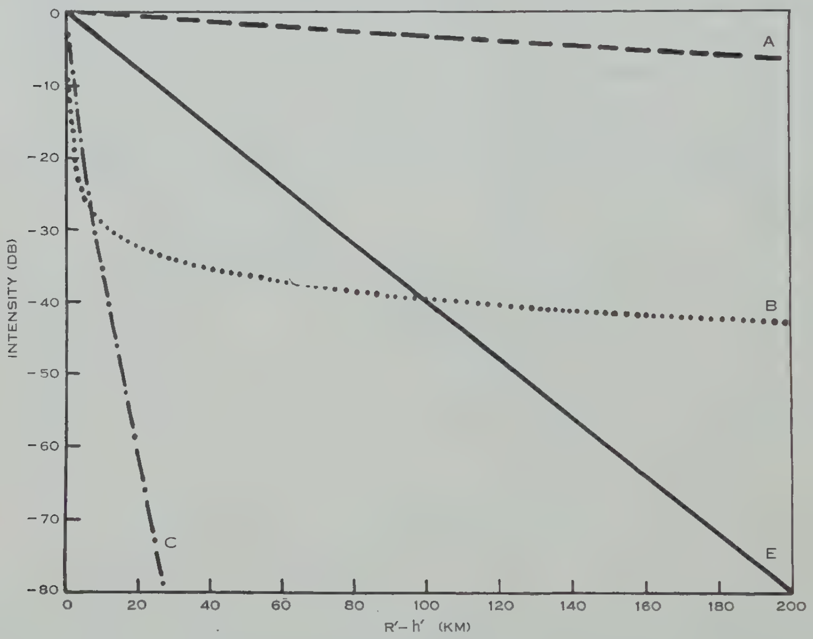


Fig. 8.—Intensity *v.* range-increment curves: *E*, empirical (from Fig. 14 of Part I); *A*, geometrical theory, cylinder of radius 20 km; *B*, Fresnel theory, sharp edge and optically flat surface; *C*, scattering theory, sharp edge, but rough surface.

Another factor influencing the intensity is absorption in the deviative region below reflection level. This increases with increasing obliquity. However, the absence of appreciable group retardation and the relatively small absorption usually observed for vertical propagation make it unlikely that this phenomenon plays a major role in the intensity fall-off.

If the rate of decline of intensity with increasing range were $0 \cdot 4$ dB/km, then, since, on the average, an echo of intensity more than 50 dB below its maximum level would fail to record, the extreme range separation observed should be 125 km, corresponding to a horizontal separation of about 300 km. For a speed of 200 km/hr this would make the satellite last 90 min. Although, in view of the large spread of values of decline rate and speed, no firm conclusions can be drawn, the fact that 125 km range separation and 90 min duration are both rather larger than the experimental averages suggests that another factor may be operating, to limit the durations of the satellites. This could be the limited lifetime of the irregularity itself.

The rapid decline towards the end of certain divergent satellites (cf. Curve *A* in Fig. 13 of Part I) and the decline, instead of growth, of the intensity of certain convergent satellites as they approach the main trace, could also be explained in terms of degeneration of the irregularity.

Some further light on this question is thrown by a comparison of the times of commencement of convergent, and termination of divergent, satellites at different stations. If genesis and degeneration, respectively, of the irregularity determine these times, they should be the same at both stations, while if they are determined by the distance of the irregularity from the station, these times would correspond rather to equal range separations and hence would only be equal when the irregularity is parallel to the line of the stations. The experimental data quoted in Section III (*g*) of Part I are not wholly conclusive, but the many cases of simultaneity, or near-simultaneity, support the idea of a limited irregularity life. Other cases show equal range separations, with non-simultaneity, and others again correspond in neither respect. This last is not inconsistent with the operation of the two effects postulated, for it is likely that genesis and degeneration will often occur at different times along the length of the irregularity.

From the experimental distribution of satellite durations, in conjunction with the above considerations, it seems likely that the average lifetime of the irregularities is a little greater than the average satellite duration and thus lies between 1 and 2 hr.

VI. Critical Frequency Differences

A small proportion of satellites are found to have critical frequencies different from those of the accompanying main trace. Higher and lower values have been observed about equally frequently. This is simply explained if we assume, as seems not unreasonable, that the ionosphere has a different maximum electron density on the two sides of the gap or step ; or as between an intermediate cloud and the edges of the gap. While one would expect that the critical frequencies for a particular satellite would tend to remain constant, and are found in most cases to do so, a gradual change could readily be explained by movement of the reflection point along a curved edge or cloud.

VII. Significance of the Velocity Determinations

There is *a priori* no clue as to whether the phenomenon is due to a general drift of an ionosphere upon which " fixed " irregularities are engraved, or some sort of wave motion, which may be quite independent of ionospheric movement. In favour of the former view, it may be pointed out that the strong tendency for disturbances to move westwards (rather than east) is consistent with the conclusion of Briggs and Spencer (1955) from results of other methods at night time, and the order of magnitude is similar. Martyn (1955) has pointed out that such a drift is well accounted for by the S field of daily magnetic variations.

If this view is adopted, the measured speed, in any particular case, may not represent the drift velocity, for the gap or step need not lie normal to the direction of drift. Where simultaneous satellites indicate different velocities, the gap must certainly be non-perpendicular to the drift ; in fact, if the directions of movement for the two satellites are known, the drift velocity can be determined.

An endeavour has been made, adopting this hypothesis, to determine the distribution of drift speeds. (Insufficient data are available on directions to permit statistical analysis.) Assuming that the histogram of Figure 4 of Part I, after smoothing, represents the distribution $N'(v')$ of " apparent " speeds, and assuming that all inclinations between the edge and the direction of drift are equally probable, it can be shown that the distribution of true drift speeds $N(v)$ is given by

$$N'(v') = \int \frac{N(v)\mathrm{d}v}{(v^2-v'^2)^{\frac{1}{2}}}.$$

This equation is not in general capable of explicit solution but it happens that the histogram fits fairly well an expression of the form

$$N'(v') = A \exp\left(-v'^2/\sigma_1^2\right) - B \exp\left(-v'^2/\sigma_2^2\right),$$

and that in this case an explicit solution is possible. This, however, leads for small values of v to negative values of $N(v)$. Attempts to find a plausible curve of $N(v)$ which gave $N'(v')$ fitting reasonably closely to the histogram proved unsuccessful.

An explanation roughly fitting all the facts would be that the phenomenon arises from a wave motion, propagated with a velocity of the order of 200 km/hr. through a layer which may itself be drifting westwards at a comparable rate, The waves may be of the nature discussed by Martyn (1950). A satisfactory theory of the phenomenon must provide an explanation not only of the velocities but also of the diurnal variation of occurrence of doublets and its correlation with the rise and fall of equivalent height of the F_2 layer discussed in Part I.

VIII. CONCLUSIONS

The main conclusions from the above discussions are :

(1) Most range multiplets at Brisbane are due to reflection at the rounded-off edges of gaps and steps in the contours at an approximately constant reflection level in the F_2 layer.

(2) A few multiplets may be due to E_s scatter with return path via F_2.

(3) The F_2 irregularities responsible for multiplets move horizontally with fairly uniform velocity.

(4) The motion thus determined is a wave motion rather than merely a general drift of the ionosphere.

(5) The duration of an irregularity is of the order of 1–2 hr.

IX. ACKNOWLEDGMENTS

This work forms part of the programme of the Radio Research Board and is published by permission of the Board. It has also been supported by the Research Committee of the University of Queensland. To these the authors wish to express their thanks.

X. REFERENCES

BECKMAN, B., MENZEL, W., and VILBIG, F. (1941).—*Telegr.-u. Fernsprtech.* **30** : 43.

BIBL, K., HARNISCHMACHER, E., and RAWER, K. (1955).—" The Physics of the Ionosphere." p. 113. (Physical Society : London.)

BRAMLEY, E. N., and ROSS, W. (1953).—*Proc. Roy. Soc.* A **207** : 251.

BRIGGS, B. H., and PHILLIPS, G. J. (1950).—*Proc. Phys. Soc. Lond.* B **63** : 907.

BRIGGS, B. H., and SPENCER, M. (1955).—" The Physics of the Ionosphere." p. 119. (Physical Society : London.)

DIEMINGER, W. (1951).—*Proc. Phys. Soc. Lond.* B **64** : 142.

ECKERSLEY, T. L. (1937).—*Nature* **140** : 846.

GIPPS, G. DE V. (1954).—Ph.D. Thesis, University of Queensland.

GIPPS, G. DE V., GIPPS, D. I., and VENTON, H. R. (1948).—*J. Coun. Sci. Industr. Res. Aust.* **21** : 215.

MARTYN, D. F. (1950).—*Proc. Roy. Soc.* A **201** : 216.

MARTYN, D. F. (1955).—" The Physics of the Ionosphere." p. 163. (Physical Society : London.)

MEEK, J. H. (1952).—*Nature* **169** : 327.

MUNRO, G. H. (1953).—*Proc. Roy. Soc.* A **219** : 447.

RATCLIFFE, J. A. (1955).—" The Physics of the Ionosphere." p. 88. (Physical Society : London.)

TOMAN, K. (1955).—*J. Geophys. Res.* **69** : 57.

THE MEASUREMENT OF THE THERMAL CONDUCTIVITIES OF ROCKS BY OBSERVATIONS IN BOREHOLES

By A. Beck,* J. C. Jaeger,* and G. Newstead†

[*Manuscript received January* 18, 1956]

Summary

The various possible methods of measuring thermal conductivity by observations in a borehole are discussed in the light of practical considerations. Three designs of measuring head are described and methods of reducing the observations are given. The possibility of obtaining results by heating a considerable length of a borehole is discussed.

I. Introduction

For the determination of the geothermal flux and similar measurements a knowledge of the thermal conductivities of rocks is necessary. In most cases these have been determined in the laboratory by the use of a steady state apparatus of the " comparison " or " divided bar " type. Such methods can easily give an accuracy of 1 per cent. for an individual specimen, but it is doubtful how far this represents the conductivity of the rock in bulk. For example measurements, made by a divided bar apparatus (cf. Benfield (1939), three and sometimes four disks were used) on 20 samples from a mass of petrologically very uniform coarse-grained granite, gave values varying apparently at random from 0·0069‡ to 0·0101 with a mean of 0·0082; this was confirmed by measurement of an additional 50 samples in a comparison apparatus (Beck 1955). These variations are probably caused in part by the fact that the diameters of the individual mineral grains (or aggregates of grains) are greater than the thicknesses of at least the thinner disks used, but they probably also represent to some extent random fluctuations in conductivity. It is difficult to know how to use these results in the calculation of heat flux, for it is clear that the association of a measured temperature with a single measured conductivity at (say) 50 ft intervals may not lead to representative results. Probably the best that can be done is to use the simple mean value, and this suggests that a method for measuring conductivities in bulk which was accurate to within 5 per cent. or even 10 per cent. would be more useful than the high precision laboratory methods. There are other important reasons for preferring measurements of rock conductivity *in situ* to laboratory determinations: firstly, the latter cannot take into account the effect of veins or open joints in the rock which may well be important, and, secondly, they are selective both in the sense that disks

---

* Australian National University, Canberra, A.C.T.

† University of Tasmania, Hobart.

‡ Units used throughout this paper are c.g.s., calorie, and °C.

can only be cut from fairly sound rock and that core may not be recovered at all from highly altered or sheared zones. The last is probably the most important consideration of all, for example as remarked in Section IV, a conductivity of 0·0042 was measured *in situ* in a highly altered and vesicular basalt, the massive portions of which (constituting the bulk of the core recovered) gave a value of 0·0052 in the laboratory.

It is clear that the measurement of thermal conductivity of rock *in situ* from a borehole would have many advantages: it would enable the conductivity to be measured at the same time as the rock temperature; it would give a result representing the average conductivity of the rock over a region determined by the length of the apparatus and the depth of penetration of heat into the rock; and, finally, it would avoid the tedious preparation of disks for use in the divided bar apparatus.

The obvious method of making such experiments is by lowering into the borehole a long probe containing a temperature measuring element and a heater which acts as a line source for supply of heat to the surrounding rock. This system has frequently been used in laboratory measurements (e.g. van der Held, Hardebol, and Kalshoven 1953; Maxwell, personal communication) and has also been used by Blackwell (1954) for measurements in short holes in tunnel walls.

The object of this paper is to discuss the possible methods of measuring rock conductivity from a single hole. Practical requirements of any method are stated in Section II. The apparatus used in field experiments is described in Section III. The results of some probe measurements are analysed in Section IV. The possibility of measuring conductivity by suddenly heating the contents of the hole is discussed in Section V, and, finally, the heating of the whole length of a hole is discussed in Section VI.

II. Experimental Conditions

Almost all the discussion of the present paper refers to field conditions: there would be little point in using apparatus of the present dimensions in the laboratory, and there is no point in developing methods which may be applicable to laboratory conditions but which cannot be used in the field. The requirements on a method are listed below.

(*a*) *Diameter of Holes*

The smallest drill commonly used for exploratory diamond drilling has a diameter of $1\frac{1}{2}$ in., which is also approximately the size of the usual jackhammer hole drilled in a tunnel wall. The apparatus described below was built for use in holes of this diameter since these are most common in Tasmania where the field work was carried out. In many holes the diameter varies, larger drill diameters being used in the upper parts of the hole, so that it is essential that it be possible to modify the apparatus in the field to take account of this; also, towards the top of a hole the diameter is usually a little greater than that of the drill and the apparatus should be sufficiently flexible to adapt itself to small changes of this sort. In addition, holes are almost always irregular, frequently there are small shearing movements across them after drilling, and occasionally loose pieces of material protrude into the hole. It is thus essential that, to provide clearance, the diameter of the measuring head should be some millimetres less than the rated diameter of the hole.

(*b*) *Contents of Holes*

Most vertical boreholes are filled with water, but occasionally a hole is encountered which is dry for portion of its length. The presence of water implies that an extremely good seal is necessary at the ends of the heating element to prevent loss of heat by convection. Experiments in glass tubes indicate that, at the rate of heat supply used, convection will be set up through a gap of 0·5 mm or even less. Holes which are not filled with water have other disadvantages, notably the high thermal resistance of the air between the heater and the rock, and the possibility of evaporation from the surface of the rock (which is usually damp) and of movement of moisture within the rock.

(*c*) *Time of Experiment*

Half an hour has been regarded as a reasonable time to be aimed at for a single measurement. If longer times are used the task of logging a deep hole becomes intolerably tedious. Also, it was felt that the danger of heat loss by convection with the seal of Figure 1 (*a*) would be greater for long runs.

The choice of this relatively short time of experiment rules out the possibility of using the approximations to the theoretical values of the temperature which are valid for large values of the time. Longer times of experiment in water-filled holes have the advantage that the depth of penetration of heat is greater, so that a more representative value of the conductivity should be obtained. For damp holes, longer times of experiment have the same disadvantage as steady state methods, namely that moisture movement and evaporation at the surface may have important and unknown effects.

III. EXPERIMENTAL MEASURING HEADS

As remarked above, the core of the problem is the provision of an adequate seal at the ends of the heating element in order to prevent convection. Three different measuring heads have been constructed which have proved satisfactory and are shown schematically in Figure 1.

The head of Figure 1 (*a*) consists of a brass rod $1\frac{1}{4}$ in. diameter and 3 ft long wound with the heating wire on its external surface. Temperature is measured by a thermistor T inside the rod, the internal arrangements being similar to that of the temperature measuring head shown in Newstead and Beck (1953). Measurement is by an A.C. bridge connected to the thermistor by a coaxial cable, the sheath of this and the supporting wire forming the heater leads L. The seals R at the ends are composed of a large number of thin rubber sheets cut radially to form a " bottle brush " ; these are attached to a poorly conducting " Perspex " cylinder and can easily be changed if holes of different diameters have to be logged. They adapt themselves very well to small changes of shape and diameter and have been observed to inhibit convection completely when the head has been operated in a glass tube. In one or two cases theoretically impossible temperature curves have been obtained which have been attributed to convection occurring through an open joint that was impossible to seal.

The head of Figure 1 (*b*) is similar to that of Figure 1 (*a*) except in length (5 ft) and in that the seals *D* are pneumatic. They consist of 8 in. lengths of inflatable rubber tube (" Ductube ") connected to an air-line hose *A* in which the heater and temperature measuring leads are carried. The " Ductubes " can be unscrewed and changed to different sizes when necessary. This arrangement will not merely give a seal which prevents convection but will completely stop flow through the hole under a considerable head. It has been used for measuring water flow in and across boreholes and will be described in detail elsewhere. The disadvantage of this head is that the air-line hose and gas cylinder for inflating it are cumbersome and that the depths to which it can be used are limited ; 300 ft has been used without difficulty.

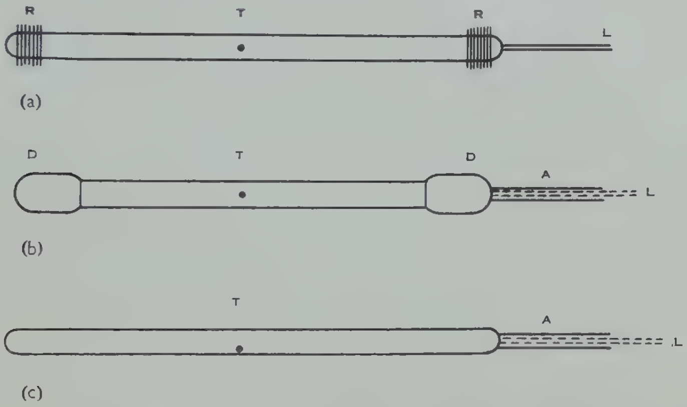


Fig. 1.—Schematic diagrams of measuring heads.

The head of Figure 1 (*c*) is an obvious development of that in Figure 1 (*b*). It consists of a rubber tube on which the heater wires are laid longitudinally, covered with thin sheet rubber, and the whole vulcanized. There are 50 wires around the circumference. The thermistor is attached to the inside of the tube. This head has the great advantage of completely eliminating the air or water space between the heater and the surface of the hole. It has been used in auger holes for measurements in soils and is very suitable for measurements in short drill holes. In such cases it can be inflated by a hand pump. As remarked earlier, in air-filled holes a gap of more than 1 mm is undesirable though gaps as large as 3 or 4 mm can be used. This implies that with any other design of head a number of heads with different diameters must be constructed if holes are at all irregular. In water-filled holes the situation is not so bad since substantial transfer of heat by convection takes place in a large water space.

It should be mentioned that in the heads of Figures 1 (*a*) and 1 (*b*) the thermistor was separated from the wall of the tube by an air- (or oil-) filled space. This was done in order to minimize the possibility of thermistors being broken or changing their constants by mechanical shocks during lowering or raising. This has had the effect of introducing a time-lag into the thermistor readings. This has been approximately corrected for as follows: if M' and c' are the mass and specific heat of the thermistor, R is the thermal resistance between it and the surrounding region, and V and v are the temperatures of this region and the thermistor, then

$$M'c'R\ \mathrm{d}v/\mathrm{d}t+v=V. \qquad (1)$$

The "time constant" $M'c'R$ of the system is measured by immersing the head in water and is of the order of $\frac{1}{2}$ min. All experimental readings are then corrected by (**1**), but only those for times greater than 3 or 4 min, where this correction is small, have been used.

IV. DISCUSSION OF EXPERIMENTAL RESULTS

The necessary theory has been given in a companion paper (Jaeger 1956*a*) which will be referred to as paper I. The system may be idealized as follows: the rock outside the hole of radius a has thermal conductivity K, density ρ, specific heat c, and diffusivity $\varkappa=K/\rho c$: the hole contains perfect conductor of thermal capacity S per unit length, to which heat is supplied at the rate Q per unit time per unit length, and there is a thermal contact resistance $1/H$ per unit area at $r=a$.

Then the temperature rise V of the perfect conductor at time t after switching on the current is given by

$$V=(Q/K)G(h,\alpha,\tau), \qquad (2)$$

where G is a complicated integral for which numerical values are given in paper I, and h, α, τ are the dimensionless parameters:

$$h=K/aH, \quad \alpha=2\pi a^2\rho c/S, \quad \tau=\varkappa t/a^2. \qquad (3)$$

The order of magnitude of these parameters is of great importance. For a hole of radius 2 cm in rock of diffusivity $\varkappa=0{\cdot}01$, $\tau=0{\cdot}3$ after 2 min and is $4{\cdot}5$ after $\frac{1}{2}$ hr: this is just the range in which neither the approximate solutions for small nor large values of the time (paper I, equations (**21**), (**22**), (**23**)) apply. Even if two terms of the asymptotic expansion for large values of the time are used, as is done by Blackwell (1954), larger values of the time are needed, particularly for large values of h.

The values of h to be expected may be seen from the fact that $h=5$ for a film of air 1 mm thick in a hole of diameter 4 cm in rock of conductivity $0{\cdot}0053$. Thus, in a dry hole values of h greater than this are to be expected. For a water-filled hole values will be correspondingly smaller. In both air- and water-filled holes some of the transport of heat through the region between the heater and the rock will be by conduction and some by convection, and, since this latter will be non-linear, the theory can only be regarded as approximate: the fact that the experimental curves agree so well with the theoretical ones indicates that it is a good approximation.

For a hole completely filled with water, α is about 1; for a hole completely filled with brass, it is about $1{\cdot}5$; if the head is not solid but tubular, higher values of α occur, for example, for the head of Figure 1 (*c*), α is about 6. In water-filled holes there is some uncertainty as to whether the whole of the thermal capacity of the water should be included in S: this implies that, even if ρc is measured in the laboratory, the value of α is not precisely known.

In principle, all of h, α, and $\varkappa$ have to be found from a single experimental curve. Blackwell proposes to find h from an approximate solution valid for small values of the time (paper I, equation (**21**)) and α and $\varkappa$ from the asymptotic expansion (paper I, equation (**23**)) for large values of the time. In the present work, using only moderate values of the time, a method of curve-fitting has usually been employed.

It follows from (**2**) that

$$\log V = \log (Q/K) + \log G(h, \alpha, \tau). \quad \ldots\ldots\ldots\ldots \quad (\mathbf{4})$$

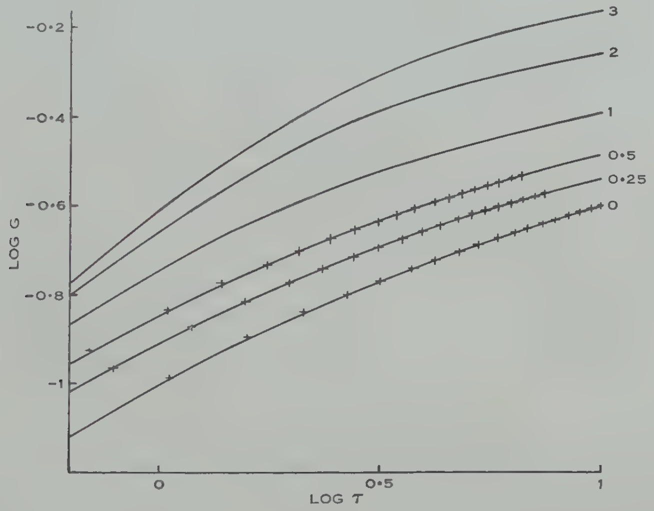


Fig. 2.—Values of log $G(h, 2, \tau)$. The numbers on the curves are the values of h. Crosses denote experimental points fitted to the curves.

A plot of $\log G(h, \alpha, \tau)$ against $\log \tau$ for $\alpha = 2$ and various values of h is shown in Figure 2. The experimental values of $\log V$ are plotted against $\log t$ on transparent graph paper and slid over a family of curves such as that of Figure 2, keeping the axes parallel, until the best fit is obtained. The displacement of the axes then gives $\log (Q/K)$ and $\log (\varkappa/a^2)$.

When this process is carried out it is found that reasonable fits may be obtained with curves of different parameters but that all these give similar values of K though they may give widely different values of ρc. Since ρc does not vary widely between rocks it is easy to make a best-possible choice of K.

As an example, a number of experimental points from a run of 100 min duration* with the head of Figure 1 (*b*) in a water-filled hole in vesicular basalt are shown by crosses in Figure 2 fitted to the different curves of the family. The values of K and ρc found from these and other curves are set out in Table 1.

---

* A number of longer runs such as this was made to check the fit of the experimental curves over a wider range.

Of these, the fit of the points with the curve $\alpha=2$, $h=0\cdot25$ is slightly the better and a rather high value such as $\rho c=0\cdot79$ might well be due to the presence of water filling the vesicles. In any case it is clear that K lies between 0·0040 and 0·0042. The present method of reduction is not regarded as being adequate to determine ρc: the value of the latter is simply used to discriminate between the various possible values of K. There are, of course, many other ways in which the experimental curves may be treated. In particular, it greatly simplifies the reduction if the temperatures for small values of the time can be relied on to be completely accurate. If this is the case, H can be determined from the relation

$$2\pi aHV=Q-S\ \mathrm{d}V/\mathrm{d}t, \qquad \ldots\ldots\ldots\ldots\ldots\ (5)$$

which holds until the change in surface temperature of the rock becomes important, and, from H and a trial value of K, h can be found. The curve of Figure 2 which has to be fitted is then known.

Table 1

POSSIBLE FITS OF EXPERIMENTAL VALUES

Curve	K	$\varkappa$	ρc
$\alpha=2$, $h=0$ ..	0·0040	0·0073	0·55
$\alpha=2$, $h=0\cdot25$ ..	0·0042	0·0053	0·79
$\alpha=2$, $h=0\cdot5$ ..	0·0046	0·0047	0·98
$\alpha=1\cdot5$, $h=0$ ..	0·0042	0·0094	0·44

Unfortunately no sound holes were available, so that conductivities measured by this method could not be compared with those found in the laboratory using core samples. For example, the basalt core from the hole in which the run above was taken was mostly so highly vesicular that its conductivity could not be measured in the laboratory, while massive samples, which were totally unrepresentative, gave a value of 0·0052.

In another hole in hard shale conductivities of 0·0079 and 0·0080 were measured below the water-table and 0·0076 above it. Laboratory measurements on surface rock (no core was available) were of this order.

A number of laboratory experiments was made in a large cement cylinder using the head of Figure 1 (a). Three separate experiments with the hole filled with water gave 0·0054, 0·0050, 0·0050, while, with the hole filled with air but the cement saturated with water, the measured values were 0·0050 and 0·0047. Conductivities for cement from the cylinder measured in the divided bar apparatus were 0·0042 for the cement dry, and 0·0050 when saturated with water.

Finally, it may be mentioned that, in the first series of field tests made, the heater was supplied from the alternator mentioned in Section VI. It was found that it was impossible to measure temperatures accurately because of cross-talk between the A.C. supplies to the heater and bridge. A regime was then adopted of running the heater on for 4 min and off for 1 min, the temperature being measured at the end of the "off" minute. There is no difficulty in principle in using such results since curves similar to those of Figure 2 can be calculated for "on-off" heating. Unfortunately, the time-lag in the thermistor made these experiments too difficult to interpret, but they are mentioned as confirming the observations of a later series (with the heater run from batteries) that the temperature curves are reproducible in successive runs in the same position to a few hundredths of a degree.

V. The Interior of the Hole Suddenly Heated

This method has been discussed by Bullard (1954) in connexion with temperature measurements by the deep-sea probe. He pointed out the possibility of measuring thermal conductivity by heating a probe rapidly by a high current.

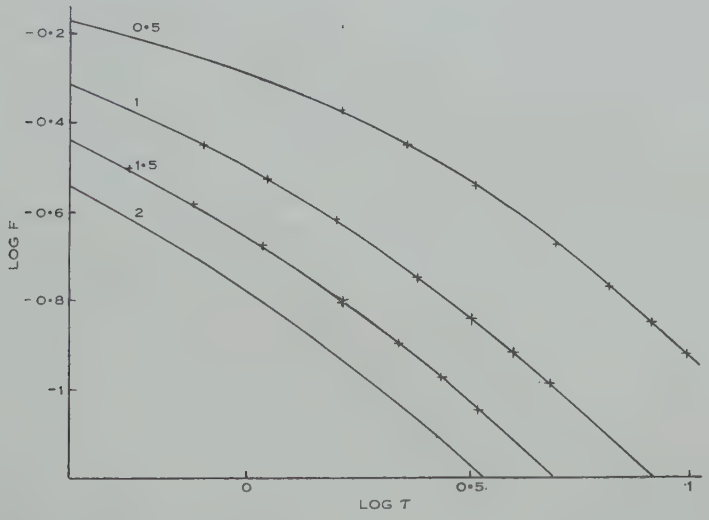


Fig. 3.—Values of $\log F(0, \alpha, \tau)$. The numbers on the curves are values of α. Crosses denote experimental points fitted to the curves.

An alternative, which is attractive because of its simplicity, is suggested by a common method of measuring temperatures in tunnel walls, namely, filling an inclined drill hole with water and measuring the temperature of the water some hours later when it has attained equilibrium with the rock. Clearly, if the water is kept well stirred and the variation of its temperature with time is measured, K can be found. Subject to the obvious practical difficulties, the method has been found to work perfectly in laboratory experiments. In this case the contact resistance may be taken to be zero, and if V_0 is the initial value of the excess of the temperature of the water above the rock and V its value at time t,

$$V/V_0 = F(0, \alpha, \tau), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

where $F(0, \alpha, \tau)$ is given in paper I, Table 1, and α and τ are defined in (3). A plot of $\log F(0, \alpha, \tau)$ against $\log \tau$ is shown in Figure 3.

To reduce the experimental results, the logarithms of the observed values of the temperature excess are plotted against log t on transparent graph paper and slid across the family of Figure 3, keeping the axes parallel, until a best fit is obtained. The displacements of the axes then give log $(\varkappa/a^2)$ and log V_0; since some heat is usually lost in adding the water it is better to regard V_0 as unknown—if it is known accurately, translation in one direction only is necessary.

In Figure 3, experimental points for the cement cylinder are shown fitted to the curves for $\alpha=0\cdot5$, $1\cdot0$, and $1\cdot5$. It may be anticipated that α is nearer to 1 than the other values and this curve gives a slightly better fit, but all curves give the same value $K=0\cdot0052$. The result by the method of Section IV was $0\cdot0051$ and in the divided bar apparatus $0\cdot0050$.

VI. The Heating of a Whole Borehole

As remarked earlier, experiments with a heating probe for times of the order of an hour, give an indication of the conductivity only in a thin region surrounding the hole (at the end of an hour in a hole of radius 2 cm the temperature at a distance of 8 cm from the surface is less than one-fifth of the surface temperature) and this is not adequate to indicate the effect of large-scale features such as joints. If experiments are carried on for much longer times very long probes must be used to ensure radial flow of heat and also the time involved in measurements at several points would be prohibitive.

One simple method of avoiding these difficulties is to heat a considerable length of the hole for a long time and to observe the temperatures at various depths, preferably with a string of thermistors as used by Balsley (personal communication). If this is done, convection will take place throughout the length of the hole and the effect of this can only be determined by experiment. It might be hoped that a stable temperature gradient would be established fairly rapidly so that the previous theory could be applied subsequently, but at the worst, temperature measurements at any depth give a complete temperature-time curve at that depth, from which the temperature distribution within the rock at the instant of switching off the current can be calculated (Jaeger 1956*b*), and this, together with the observed temperatures in the hole after switching off, gives an integral equation, which can be solved numerically for K.

In view of this, and also of the possibility of obtaining information about the time which must elapse after drilling for a borehole to attain equilibrium, it was decided to try the experiment of heating a length of a borehole. To provide power, a 3 kW alternator was mounted in a " Landrover ". Unfortunately, on examination, the only suitable hole had a small water flow from it, but as it was felt that this might be superficial, the experiment was continued. A length of 285 ft of cable was lowered into the hole, the power to the heater being 2400 W or $0\cdot066$ cal cm^{-1} sec^{-1}. The hole was heated for 700 min, the current being then switched off and the recovery of temperature observed. Temperature measuring heads were located at depths of 75 and 175 ft, their records being shown in Figure 4. The hole was in dolerite whose average conductivity as measured in the laboratory was $0\cdot0048$ and its diameter was $2\frac{5}{16}$ in. from the collar to 128 ft, and $1\frac{7}{8}$ in. below that.

The dotted curves are calculated as in Section IV on the assumption that there is no movement of water. As remarked earlier, there will certainly be convection in the early stages of heating but this is not sufficient to account for the differences between the observed and calculated curves since this difference is also shown by the cooling curves, in which the effect of convection should be absent. Curves tending to a horizontal asymptote were shown in paper I, Section VI, to be characteristic of heat exchange by movement of water in the hole, and a small flow of water up the hole undoubtedly accounts for the present results (this flow must be much smaller than that issuing from the collar of the hole, which must be a surface effect).

These results, particularly the nature of the cooling curves, indicate that the method is well worth trying in a hole in which there is no water movement.

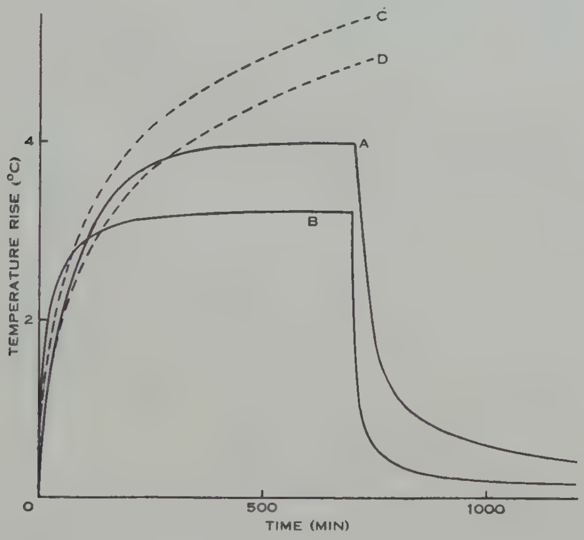


Fig. 4.—Temperatures at 75 ft (curve *A*) and 175 ft (curve *B*) in a borehole heated to 285 ft. Dotted curves *C*, *D* are those expected in the complete absence of water movement.

VII. Movement of Water Along and Across Boreholes

One by-product of the present experiments was the discovery that water movement in boreholes occurs very frequently. It has occasionally been suspected and invoked to explain anomalous temperature gradients, but apparatus of the present type is an extremely sensitive detector of it: the use of the head of Figure 1 (*b*) to measure water flows in or across boreholes will be discussed elsewhere since these measurements are of considerable engineering importance.

In the present series of experiments in the Tasmanian dolerites and basalts, water flow was detected in every hole investigated. In one hole in particular, which was apparently quite normal, being full to the collar and having no water flowing from it, water entered at 400 ft, flowed up the hole and out at 50 ft.

VIII. ACKNOWLEDGMENTS

We are indebted to the Hydro-electric Commission of Tasmania and the Snowy Mountains Hydro-electric Authority for permission to use their boreholes and for putting their facilities at our disposal.

IX. REFERENCES

BECK, A. (1955).—Temperature and conductivity measurements in the Snowy Mountains Region. Dep. Rep., Aust. Nat. Univ.

BENFIELD, A. E. (1939).—*Proc. Roy. Soc.* A **173** : 428.

BLACKWELL, J. H. (1954).—*J. Appl. Phys.* **25** : 137.

BULLARD, E. (1954).—*Proc. Roy. Soc.* A **222** : 408.

VAN DER HELD, E. F. M., HARDEBOL, J., and KALSHOVEN, J. (1953).—*Physica, 's Grav.* **19** : 208.

JAEGER, J. C. (1956*a*).—*Aust. J. Phys.* **9** : 167.

JAEGER, J. C. (1956*b*).—*J. Math. Phys.* **34** : 316.

NEWSTEAD, G., and BECK, A. (1953).—*Aust. J. Phys.* **6** : 480.

CORRIGENDUM

VOLUME 8, NUMBER 4, PAGE 565

"*The Angular Size of the Variable Radio Source Hydra–A*"

The average central brightness temperature given on page 565 is incorrect and should be 3×10^6 °K. No other part of this paper is affected by this change.

TWO-DIMENSIONAL AERIAL SMOOTHING IN RADIO ASTRONOMY

By R. N. Bracewell*

[*Manuscript received January* 27, 1956]

Summary

The *visibility* of a Fourier component of a two-dimensional temperature distribution which is scanned by certain kinds of rigid aerial is given by the normalized complex autocorrelation function of the field distribution over the aerial aperture (assuming that turning the aerial in its own plane is not allowed). Hence, for finite aerials, the visibility of the Fourier components falls to zero at finite values of spatial frequency. Consequently observations need only be made at certain *peculiar intervals* whose size is worked out. *Interpolation* between observations so spaced can be carried out by a method which then, by a simple extension, permits *filtering* of data which are to be freed from high spatial frequencies. Both interpolation and filtering are basic processes in the handling of two-dimensional data and contour maps in radio astronomy. The *restoration* of smoothed data is discussed from the viewpoint that only the simplest operations on extensive two-dimensional data are feasible, and details of a suitable technique of restoration are summarized. Application of *further smoothing* to existing data is shown to be important, and a method for doing it is given, again under the restriction to simple operations. The *flux density* of a source is shown to be given exactly by summing one in four of the isolated values observed at the peculiar intervals.

I. Introduction

The essential mathematical phenomena of aerial smoothing have been set out by Bracewell and Roberts (1954) in a study (henceforth referred to as paper I) of the one-dimensional case ; the aim of the present paper is to give the generalization to two dimensions. On advancing from one to two dimensions, one meets a new phenomenon which is connected with the rotation of celestial objects in the field of view of some optical instruments (e.g. sextants, theodolites, siderostats) but not others (equatorially mounted telescopes, coelostats). Identical aerials pointed in the same direction in space, but mounted differently, may " see " different temperatures as a result of their " fields of view " being differently oriented. Thus, in specifying how an aerial beam is pointed, one must give the two direction coordinates and also the " position angle " ψ. Section II of this paper develops the general equations which form the basis of two-dimensional aerial smoothing but the remainder of the paper is restricted to the case where ψ is constant over the sky. When ψ is not constant we obtain a branch of the theory which includes the important problem of strip integration, a development which is covered in another paper (Bracewell 1956).

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney ; present address : Stanford University, Stanford, California.

If we suppose a two-dimensional temperature distribution $T(x,y)$ analysed into components of the form

$$\begin{matrix}\cos\\ \sin\end{matrix}\,(2\pi ux-\varphi_u)\;\begin{matrix}\cos\\ \sin\end{matrix}\,(2\pi vy-\varphi_v),$$

then the effect of aerial smoothing is to eliminate the component if u or v exceeds certain limits, and to modify its strength relative to other components if it is not eliminated. On the uv-plane the points representing eliminated components lie outside a central area, usually circular or rectangular. This is proved in Section III and the consequences are worked out in Section IV, which establishes the discrete interval theorem according to which an observed distribution is fully determined by spot values lying on a rectangular lattice.

which Ewing

Section V discusses the construction of contour diagrams from data spaced at discrete intervals (interpolation) and a procedure for ensuring that contour diagrams are free from detail which is not warranted.

In Section VI the question of correcting for aerial smoothing is taken up and emphasis is laid on the necessity for simplicity in any operation that is proposed for application to an extensive two-dimensional array of data.

Section VII deals with the application of aerial smoothing, an operation which is called for in the reduction of data. Again the approach is from the direction of simple operations. In Section VIII it is shown that the integral over a true temperature distribution is given correctly by the integral over the observed distribution and that this integral is exactly evaluated by summing only one in four of the spot values just sufficient to determine the distribution.

II. THE BASIC EQUATION

Let $T(\theta, \varphi)$ be the distribution of brightness temperature for the frequency and polarization accepted by the aerial, θ and φ being the co-declination and Right Ascension respectively.

To specify the orientation of the aerial it is necessary to give not only (θ_0, φ_0), the co-declination and Right Ascension towards which the aerial is pointed, but also a position angle which determines the rotation of the aerial about this main axis. The position angle ψ gives the direction of a transverse axis, fixed in the aerial, measured eastwards from north (Fig. 1).

$A_1(\alpha, \beta)$ is the directional diagram of the aerial, α and β being spherical polar coordinates relative to axes fixed in the aerial. The polar angle α is measured from the main axis; β from the great circle containing the main and transverse axes. $A_1(\alpha, \beta)$ is supposed normalized so that

$$\int_0^{2\pi}\int_0^{\pi} A_1(\alpha, \beta)\sin\alpha\, d\alpha d\beta=1. \quad \ldots\ldots\ldots\ldots \quad (1)$$

When the aerial is pointed towards (θ_0, φ_0) with position angle ψ, the temperature measured is a weighted mean of $T(\theta, \varphi)$:

$$T_a(\theta_0, \varphi_0, \psi)=\int_0^{2\pi}\int_0^{\pi} A_1(\alpha, \beta)T(\theta, \varphi)\sin\alpha\, d\alpha d\beta. \quad \ldots\ldots \quad (2)$$

The co-declination and Right Ascension (θ, φ) appearing in the integral are related to α, β, θ_0, φ_0, and ψ as in the spherical triangle of Figure 1, from which we have the relations

$$\cos\theta = \cos\theta_0 \cos\alpha + \sin\theta_0 \sin\alpha \cos(\beta-\psi),$$
$$\cot(\varphi_0-\varphi) = -\cos\theta_0 \cot(\beta-\psi) + \sin\theta_0 \cot\alpha \operatorname{cosec}(\beta-\psi).$$

The basic formula does not appear to have been given explicitly before. Among the conditions for its validity are two of interest. Firstly, it has been assumed that powers from different directions add, which would not be the case if there were any coherence in phase. Such coherence can arise from ground reflection and it is actually invoked in the case of the sea interferometer, to which the present theory is not directly applicable. Coherence between radiations from different parts of the sky have so far been detected only as atmospheric scintillations. Secondly, it should be noted that the equation applies only to rigid aerials for which the aerial pattern does not change with different positions of the aerial. Interferometers whose parts may be tilted or rotated form instances of deformable aerials, as do equatorially mounted aerials which receive any radiation (via the main beam or side lobes) from the ground.

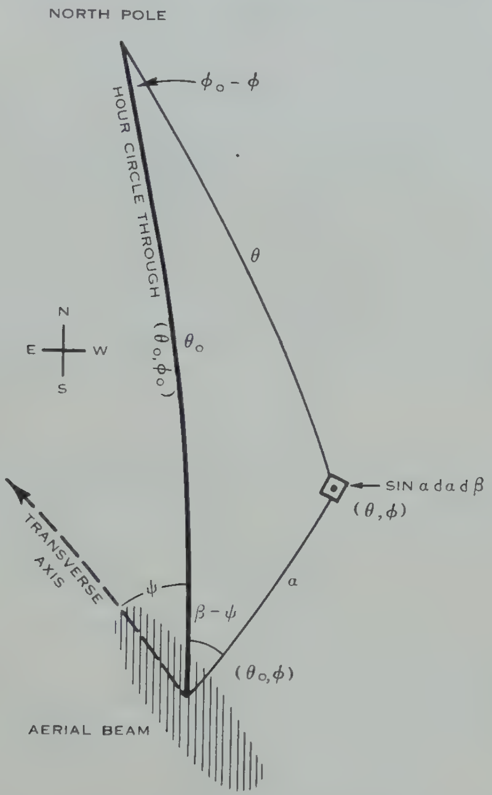


Fig. 1.—Spherical triangle showing the relation between the celestial coordinates θ, φ and the aerial coordinates α, β, ψ, θ_0, φ_0.

For the present purposes it is desirable to specialize equation (**2**) to a simpler form which still includes essential phenomena of two-dimensional smoothing. To do this we note that $\theta-\theta_0$ and $\varphi-\varphi_0$ will be approximately rectangular coordinates if attention is limited to a small fraction of the celestial sphere (not near the poles). Let x and y be the approximate rectangular coordinates of the point at which the aerial is directed, assume ψ to be constant, and use the fact that with a sufficiently directional aerial the area of sky contributing to the received power is so minute that the limits of integration are unimportant and may for convenience be replaced by infinite limits. Equation (**2**) can then be replaced by one of the form

$$T_a(x, y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A(x-\xi,\ y-\eta)T(\xi, \eta)\mathrm{d}\xi\mathrm{d}\eta, \quad \ldots\ldots\ldots\ldots \quad (\mathbf{3})$$

where ξ and η are the coordinates of the element $\mathrm{d}\xi\mathrm{d}\eta$, and $T(x, y)$ and $T_a(x, y)$ are respectively the true and measured brightness temperatures at (x, y). The function $A(\xi-x,\ \eta-y)$ representing the directional diagram is related to the definitive directional pattern $A_1(\alpha, \beta)$ in a way which involves the declination of the area where the approximate rectangular coordinates are to be drawn; therefore, in all that follows, the further assumption is implied that attention is confined within a zone of declination (or other appropriate coordinate, such as galactic latitude) narrow enough for the function A to have the same form throughout.

III. THE SPECTRAL VISIBILITY THEOREM

The two-dimensional Fourier transform of $T_a(x, y)$ is defined by

$$\bar{T}_a(u, v)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}2\pi(ux+vy)}T_a(x, y)\mathrm{d}x\mathrm{d}y. \quad \ldots\ldots \quad (\mathbf{4})$$

Then, by the two-dimensional convolution theorem (Sneddon 1951),

$$\bar{T}_a(u, v)=\bar{A}(u, v)\bar{T}(u, v). \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{5})$$

Equation (**3**), which takes the form of a two-dimensional convolution, we shall write as

$$T_a=A*T$$

following a convenient notation† used in paper I.

The important quantity $\bar{A}$, which will be called the spectral visibility function, has a simple relationship to the aerial aperture distribution of electric field $E(x/\lambda, y/\lambda)$, where x and y are rectangular coordinates in the plane of the aperture. We know that a two-dimensional aperture distribution and the angular spectrum of its accompanying field, P ($\sin\gamma$, $\sin\delta$) are Fourier transforms one of the other, provided we adopt as independent variables x/λ, y/λ,

---

† For the sake of simplicity it will be assumed that $A(x,y)$ is symmetrical about both the x–and y–axes. The asterisk notation for convolution is then not complicated by the need to draw a distinction between " aerial directional diagram " and " response to a point source ".

$\sin\gamma$, $\sin\delta$, where the angles γ and δ are measured from the planes, normal to the plane of the aperture, which contain respectively the y- and x-axis. Thus

$$P\ (\sin\gamma,\ \sin\delta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\left(\frac{x}{\lambda},\frac{y}{\lambda}\right)\exp\left[\mathrm{i}2\pi\left(\frac{x}{\lambda}\sin\gamma+\frac{y}{\lambda}\sin\delta\right)\right]\mathrm{d}\left(\frac{x}{\lambda}\right)\mathrm{d}\left(\frac{y}{\lambda}\right),$$

and

$$E\left(\frac{x}{\lambda},\frac{y}{\lambda}\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P\ (\sin\gamma,\ \sin\delta)\ \exp\left[-\mathrm{i}2\pi\left(\frac{x}{\lambda}\sin\gamma+\frac{y}{\lambda}\sin\delta\right)\right]$$
$$\times\,\mathrm{d}\ (\sin\gamma)\mathrm{d}\ (\sin\delta).$$

We may write this latter equation

$$E=\bar{P},$$

where the bar indicates, here and below, the two-dimensional Fourier transform. For real values of γ and δ, P is proportional to the field directional diagram. Hence the power directional diagram A is given by

$$A=\text{const.}\ |\,P\,|^2,$$

or if P is not complex, by

$$A=\text{const.}\ P^2.$$

Hence, by the convolution theorem,

$$\begin{aligned}\bar{A}&=\text{const.}\ \bar{P}*\bar{P},\\ &=\text{const.}\ E*E.\qquad\qquad (6)\end{aligned}$$

Thus the spectral visibility function $\bar{A}$ is proportional to the self convolution of the electric field distribution.† The constant of proportionality is fixed by the requirement that $\bar{A}$ (0) should be unity, a consequence of the normalization of A. Finally we have

$$\bar{T}_a=\text{const.}\ (E*E)\bar{T}.\qquad\qquad (7)$$

The spectral visibility theorem (7) is useful both as a mental aid to seeing the general form of $\bar{A}$ and also as a means for computing it.

A most important property of $\bar{A}$ is that, for aerial apertures of finite extent, it falls to zero outside a central region in the uv-plane. The shape of this region can readily be determined graphically by performing the self convolution indicated in equation (6). In general the shape is complicated, but for circular apertures it is circular, and for rectangular apertures, rectangular.

An observed distribution taken with an aerial consisting of a finite aperture will, by equation (7), have the same property as $\bar{A}$, namely, that its transform

---

† In the more general case of P complex and not symmetrical, $\bar{A}$ is the (normalized) complex autocorrelation function of E,

$$\bar{A}(u,v)=\frac{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E(x/\lambda-u,\ y/\lambda-v)E^*(x/\lambda,\ y/\lambda)\mathrm{d}(x/\lambda)\mathrm{d}(y/\lambda)}{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E(x/\lambda,\ y/\lambda)E^*(x/\lambda,\ y/\lambda)\mathrm{d}(x/\lambda)\mathrm{d}(y/\lambda)}.$$

is zero outside a central region. The consequences of this property are given in the next section.

In what follows it is occasionally convenient to say that a function is "cut off" if there exists a central region outside which it is everywhere zero. Also the term "spectrum" is occasionally used as a synonym for Fourier transform.

IV. The Discrete Interval Theorem

A function $T_a(x, y)$ such that $\bar{T}_a(u, v)$ is zero for $|u| \geqslant u_c$ or $|v| \geqslant v_c$, is completely determined by its values at the points $(m/2u'_c + a, n/2v'_c + b)$, where m and n assume all integral values, a and b are arbitrary constants, and the spacing between points may be as wide as is compatible with $u'_c \geqslant u_c$ and $v'_c \geqslant v_c$.

The condition on $\bar{T}_a(u, v)$ may be expressed by saying that it is zero on and outside a rectangle which is centred on the origin of the uv-plane and set with its sides parallel to the axes; and we may note that this covers the case of $\bar{T}_a(u, v)$ zero on and outside a circle or other region provided the rectangle is chosen sufficiently large.

It is sufficient to give a proof for the case where a and b are zero, i.e. where the origin of x and y is one of the sampling points. For, if the transform of $T_a(x, y)$ is zero on and outside a given rectangle, so also is that of $T_a(x+a, y+b)$ by virtue of the two-dimensional shift theorem, according to which the Fourier transform of $T_a(x+a, y+b)$ is $\bar{T}_a(u, v) \exp\{-i2\pi(au+bv)\}$, which must be zero where $\bar{T}_a(u, v)$ is zero. Therefore, if the theorem is true for $T_a(x, y)$, it is also true for $T_a(x+a, y+b)$; but values of $T_a(x+a, y+b)$ at points of an array which includes the origin are values of $T_a(x, y)$ taken over an offset array.

To prove the theorem we use the bed-of-nails function ${}^2\mathrm{III}(x, y)$ consisting of a two-dimensional array of unit impulses separated by unit distance. Thus

$$ {}^2\mathrm{III}(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} {}^2\delta(x-m, y-n), $$

where the two-dimensional impulse function ${}^2\delta(x, y) = 0$ for $x \neq 0$ or $y \neq 0$, but

$$ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} {}^2\delta(x, y)\, \mathrm{d}x \mathrm{d}y = 1. $$

The bed-of-nails function is shown in Appendix I to be its own two-dimensional Fourier transform.

Proof of Theorem

Let $\bar{F}(u, v) = (4u'_c v'_c)^{-1}\, {}^2\mathrm{III}(u/2u'_c, v/2v'_c) * \bar{T}_a$, a function which may be pictured as an array of islands in the uv-plane, each the same as $\bar{T}_a$, spaced at intervals $2u'_c$ in the u direction and $2v'_c$ in the v direction. The islands will not overlap (but may touch) if $u'_c \geqslant u_c$ and $v'_c \geqslant v_c$. Under this condition, in the central region where $|u| < u_c$ and $|v| < v_c$, we have

$$ \bar{F}(u, v) = \bar{T}_a. $$

Hence $\overline{T}_a$ may be recovered from $\overline{F}(u, v)$, and consequently T_a may be recovered from $F(x, y)$, the two-dimensional Fourier transform of $\overline{F}(u, v)$. But, by the two-dimensional convolution theorem,

$$F(x, y)=4u'_c v'_c\ {}^2\mathrm{III}(2u'_c x,\ 2v'_c y)T_a(x, y),$$

which contains values of T_a only at discrete intervals $(2u'_c)^{-1}$ and $(2v'_c)^{-1}$ of x and y. Hence T_a is completely determined by its values at discrete intervals of x and y which are equal to or less than $(2u_c)^{-1}$ and $(2v_c)^{-1}$. Since these intervals are peculiar to each aerial they will be referred to as the peculiar intervals.

V. Interpolation and Filtering

If the distribution T_a is measured at suitable discrete intervals of x and y, the numerical problem of interpolating will often arise. Let the scale factors be changed so that $(2u_c)^{-1}$ and $(2v_c)^{-1}$ both become unity. Then it is required to calculate $T_a(x, y)$ for any x and y, given ${}^2\mathrm{III}T_a$. From the preceding section we can say that

$$\overline{T}_a=\overline{M}\ \overline{{}^2\mathrm{III}T_a},$$

where

$$\overline{M}(u, v)=\begin{cases}1, & (|u|<\tfrac{1}{2} \text{ and } |v|<\tfrac{1}{2}),\\ 0, & (|u|>\tfrac{1}{2} \text{ or } |v|>\tfrac{1}{2}).\end{cases}$$

In other words, we formally express the recovery of $\overline{T}_a$ from the array of islands $\overline{F}$ (written now as $\overline{{}^2\mathrm{III}T_a}$) as the operation of multiplying by $\overline{M}$, the parallelopiped function of unit height, length, and breadth. Then, by the two-dimensional convolution theorem,

$$T_a=M*({}^2\mathrm{III}T_a). \qquad \text{(8)}$$

Now it can be shown that

$$M=\frac{\sin \pi x \sin \pi y}{\pi^2 xy},$$

but because of the character of ${}^2\mathrm{III}$ the convolution integral (8) reduces exactly to the sum of products of the array of values of T_a at discrete intervals with the array of values of M taken at the same intervals. In calculating any particular value, therefore, we need only know a discrete set of values of M. There are two principal interpolations for which it is useful to have the M array recorded for reference. The points A and B (Fig. 2) show the relation of the required points to the lattice points where T_a is known. For interpolation at A, the centre of the square, we use

$${}^2\mathrm{III}(x-\tfrac{1}{2},\ y-\tfrac{1}{2})M,$$

and for interpolation at B, the mid point of a lattice line, we use

$${}^2\mathrm{III}(x-\tfrac{1}{2},\ y)M.$$

These two arrays are given in Table 1 and Table 2. For interpolation at C we use the B array with rows and columns interchanged. The reason that interpolation at B needs only the values of T_a in the row through B is that a cross section of a surface whose (two-dimensional) spectrum is cut off, itself has a cut-off (one-dimensional) spectrum. The amount of work needed to perform A-interpolation considerably exceeds that for B-interpolation, so that in practice it will often be preferable to use only B-interpolation, proceeding by steps such as D, E, F, G (Fig. 2). Table 2 is given at greater length in Table 1 of paper I.

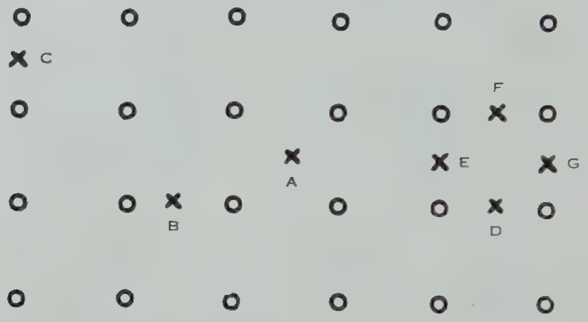


Fig. 2.—Known values of T_a are marked ○. Interpolated values are required at A, B, and C.

TABLE 1

ARRAY FOR A-INTERPOLATION

(The first octant is shown at greater length on the right)

		0·08	0·08							0·005
	0·05	$\overline{0\cdot14}$	$\overline{0\cdot14}$	0·05					0·008	$\overline{0\cdot006}$
0·08	$\overline{0\cdot14}$	0·41	0·41	$\overline{0\cdot14}$	0·08			0·016	$\overline{0\cdot011}$	0·009
0·08	$\overline{0\cdot14}$	0·41	0·41	$\overline{0\cdot14}$	0·08		0·045	$\overline{0\cdot027}$	0·019	$\overline{0\cdot015}$
	0·05	$\overline{0\cdot14}$	$\overline{0\cdot14}$	0·05		0·405	$\overline{0\cdot135}$	0·081	$\overline{0\cdot058}$	0·045
		0·08	0·08							

We have so far assumed that $\bar{T}_a$ is cut off outside the rectangle determined by $|u|=u_c$, $|v|=v_c$, without enquiring whether it is also zero anywhere within the rectangle; but if the aerial has circular symmetry, as is often the case, the cut-off boundary will be circular.

Let $\bar{A}$ fall to zero on the circle $u^2+v^2=R^2$. Then

$$\bar{T}_a=\bar{N}\ {}^2\overline{\mathrm{III}T}_a,$$

where

$$\bar{N}(u, v)=\begin{cases}1, & (u^2+v^2<R),\\ 0, & (u^2+v^2>R).\end{cases}$$

Hence

$$T_a=N*({}^2\mathrm{III}T_a),$$

and we know that

$$N=\frac{R\mathrm{J}_1\{2\pi R\sqrt{(x^2+y^2)}\}}{\sqrt{(x^2+y^2)}}.$$

TABLE 2

ARRAY FOR B-INTERPOLATION

			0	0			
		0	0	0	0		
——	0·127	$\overline{0\cdot212}$	0·637 —	— 0·637	$\overline{0\cdot212}$	0·127	——
		0	0	0	0		
			0	0			

The array for A-interpolation in this case is shown in Table 3. When the aerial is elongated and produces a " fan beam ", then as long as the position angle of the beam is kept fixed, this case is covered by the scale change making $(2u_c)^{-1}$ and $(2v_c)^{-1}$ unity. In practice a rectangular lattice of points is used to define the data.

TABLE 3

ARRAY FOR INTERPOLATION WHEN THE AERIAL HAS CIRCULAR SYMMETRY

(Only the first octant is shown)

				0·005
			0·016	$\overline{0\cdot016}$
		$\overline{0\cdot027}$	0·005	0·006
	$\overline{0\cdot025}$	0·038	$\overline{0\cdot030}$	0·022
0·391	$\overline{0\cdot102}$	0·046	$\overline{0\cdot027}$	0·018

The operation of filtering is closely related to interpolation, but the independent variables, instead of assuming discrete values, vary continuously. It is required to remove the spectral components outside a certain cut-off boundary—which is precisely the means by which interpolation was effected. However, the integral does not reduce to a summation as in that case. The filtered value T_a corresponding to an unfiltered value T is given by

$$T_a = M * T,$$

in the case of a rectangular cut-off boundary. To evaluate this integral we would calculate a summation

$$T_{\xi\eta} = \left\{ {}^2\mathrm{III}\left(\frac{x}{\xi}, \frac{y}{\eta}\right) M \right\} * T,$$

which would approach the desired integral as the spacings ξ and η approached zero. Just how small they need be we shall now enquire.

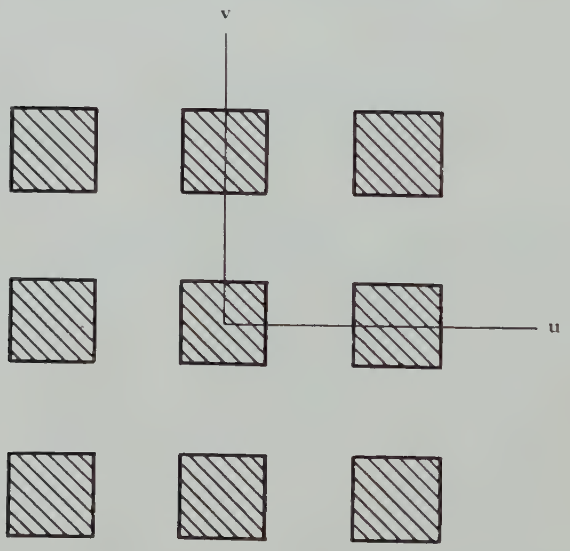


Fig. 3.—To illustrate the function $\frac{1}{4}{}^2\mathrm{III}(\frac{1}{2}u, \frac{1}{2}v) * \overline{M}$.

Beginning with the coarsest interval $\xi = \eta = 1$, we find $T_{11} = T$, that is, no filtering has been achieved; but with $\xi = \eta = \frac{1}{2}$, we find

$$T_{\frac{1}{2}\frac{1}{2}} = \{ {}^2\mathrm{III}(2x, 2y) M \} * T,$$

whence

$$\overline{T}_{\frac{1}{2}\frac{1}{2}} = \frac{1}{4}\left\{ {}^2\mathrm{III}\left(\frac{u}{2}, \frac{v}{2}\right) * \overline{M} \right\} \overline{T}.$$

Thus $\overline{T}_{\frac{1}{2}\frac{1}{2}}$ consists of a central part $\overline{M}\,\overline{T} = \overline{T}_a$ plus islands in the uv-plane covering only one-quarter of the plane, as shown in Figure 3. For many purposes this would be sufficient filtering, since the components to be rejected would often lie chiefly just beyond the central region.

The test for adequacy of filtering is to compare B-interpolations with T in the most unfavourable areas. Where further filtering is required, further application of the same process with $\xi=\eta=\frac{1}{4}$ will remove the eight peripheral islands of Figure 3 as well as others further out. Table 4 gives for permanent reference the array needed for the summations to which filtering is reduced by the suggested procedure. Subsequent stages of filtering, where indicated, use exactly the same array.

A further comment may be made on a procedure which appears at first sight to be equivalent to filtering. If T is read off at discrete intervals corresponding to the desired cut-off, the set of values so obtained defines a function with cut-off spectrum which may then be recovered by interpolation. But this function, $M*({}^2\text{III}T)$, is not the same as T_a since it is contributed to by high frequency components of T. A use for it is suggested in Section VII.

TABLE 4

ARRAY FOR FILTERING

						0
					0·016	0
				0	0	0
			0·045	0	$\overline{0\cdot027}$	0
		0	0	0	0	0
	0·101	0	$\overline{0\cdot034}$	0	0·020	0
0·250	0·159	0	$\overline{0\cdot053}$	0	0·032	0

Both interpolation and filtering have application to the handling of contour diagrams of brightness distribution. When data have been obtained at the peculiar intervals, it is not always easy to draw the contours in by eye, especially in places where the contours are tightly bunched or strongly curved. Interpolation of extra values is here useful. On the other hand, when data have been obtained continuously or semi-continuously high frequencies can appear accidentally, and may be revealed by interpolation and removed by filtering. One source of spurious high frequencies is undoubtedly the freehand interpolation of contours—the apparent detail so generated should be ignored during subsequent interpretation. In fact, before any spatial detail near the limit of resolution is discussed, it should be filtered to ensure that no spurious feature is included which is unjustified on the known aerial resolution.

It is unlikely that any of the published surveys of galactic noise have been filtered and it is certain that many of them contain unwarranted detail. It is suggested that when publishing a contour diagram one should make clear how the interpolation was done and whether filtering was done. As an alternative, or adjunct, to a contour diagram, a table of values at the peculiar interval would have merits.

VI. Restoration

The possibilities of restoration have been studied in paper I for one dimension and they are the same in the present case. In brief, nothing can be done about spectral components near or beyond cut-off, but the correct relative proportions of the remaining components can be restored. Whether a given T_a can be gainfully restored depends on $\bar{T}$—whether it falls to an unimportant level before the aerial cuts it off. If it does so, the distribution is fully resolved and may be confidently restored. More usually, however, $\bar{T}$ is such that the justifiable degree of restoration is a nice compromise between improvement due to the restored balance of the low frequency components and deterioration due to overshoot and magnification of errors.

The method of successive substitutions is applicable in two dimensions but has the disadvantage of requiring an amount of computation that renders it infeasible in much of the recent extensive high resolution work. There therefore appears to be scope for a less elaborate procedure which would at least indicate the order of magnitude of the corrections and enable one to see where fuller attention to restoration might be wise. The special case of a Gaussian aerial has been approached on this basis (Bracewell 1955*a*) with encouraging results. The method is a generalization of one worked out for the one-dimensional case (Bracewell 1955*b*) in which the correction is given by the amount by which the distribution lies above the mid point of a certain chord. In two dimensions it is useful to imagine the corresponding plane, but the calculation must be done numerically, not graphically. The amount of the correction is

$$-\tfrac{1}{4}(\Delta_{xx}+\Delta_{yy})T_a,$$

where Δ_{xx} is the second difference of T_a when y is kept constant, and the interval over which the differencing is done is equal to $\sqrt{2}$ times the standard deviation of the Gaussian beam. In the case of beams other than Gaussian the correction becomes

$$-(\chi\Delta_{xx}+\psi\Delta_{yy})T_a,$$

where χ, ψ, and the differencing intervals α and β are fixed by matching $\chi \sin^2 \pi\alpha u+\psi \sin^2 \pi\beta v$ to $\bar{A}^{-1}-1$ as previously described for the one-dimensional case (Bracewell 1955*c*).

VII. Carrying Out Aerial Smoothing

The need to carry out aerial smoothing numerically can arise in various ways :

(*a*) In comparing data obtained with different aerials but on the same wavelength, further smoothing on the better resolved data must be carried out.

(*b*) In determining spectra by comparison of data taken on different wavelengths, different resolving powers must usually be allowed for.

(*c*) In utilizing optical data presented in the form of contour diagrams or counts showing the frequency of occurrence of celestial objects, some numerical operation akin to aerial smoothing must first be applied.

One might consider restoring the less well-resolved data before making comparison, and in some cases this might be desirable, but more often the sacrifice of detail in the procedure discussed here will be more than compensated for by the reliability of the comparison; for a smoothed distribution can be uniquely determined whereas there is uncertainty in the determination of a restored distribution.

Case (*c*) differs from (*a*) and (*b*) through the presence of high frequency components resulting from the high resolution of optical data, and such data should first be roughly filtered by taking running means, smoothing by eye, or some other appropriate method.

Since the present problem has arisen in connexion with aerials having Gaussian beams, Gaussian smoothing will be used to illustrate the proposed methods.

Let

$$A(x, y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right),$$

where σ is the standard deviation of any section through $x=0$, $y=0$. Then

$$\bar{A}(u, v)=\exp\{-2\pi\sigma^2(u^2+v^2)\}.$$

We now take T to stand for a distribution which is to be smoothed and let T_a represent the result of the smoothing. The operation by which T_a is obtained from T must be such that

$$\begin{aligned}\bar{T}_a&=\bar{A}\,\bar{T}\\&=\exp\{-2\pi\sigma^2(u^2+v^2)\}\bar{T}.\end{aligned}$$

We shall take as an approximation to this exact formula, for reasons which follow later,

$$\bar{T}_a=\tfrac{1}{2}\{1+\cos 2\sqrt{(2\pi)}\sigma u\cos 2\sqrt{(2\pi)}\sigma v\}\bar{T}. \quad \ldots\ldots\ldots\ldots \quad (9)$$

The degree of agreement with the accurate Gaussian function is shown in Figure 4. Applying the Fourier transformation to equation (9) we have

$$\begin{aligned}T_a(x, y)=\tfrac{1}{2}T(x, y)+\tfrac{1}{8}\{&T(x-x_1, y-x_1)+T(x+x_1, y-x_1)\\&+T(x+x_1, y+x_1)+T(x-x_1, y+x_1)\}, \quad \ldots\ldots\ldots\ldots \quad (10)\end{aligned}$$

where

$$x_1=\sqrt{(2/\pi)}\sigma.$$

Equation (10) is interpreted diagrammatically in Figure 5 which shows a square grid ruled at intervals x_1. First take the mean of the four values surrounding the value (100) to be corrected and then average the result with 100. Alternatively, average the differences shown in smaller figures and subtract half the amount from 100. The value for T_a is thus 98·5. The possibility of reducing the numerical operations to this simple procedure of averaging was the reason for selecting the approximating function in equation (**9**).

The amount of work involved in applying the operation here described is consistent with the economy of effort desirable in any two-dimensional operation, and is much less than the work involved in carrying out smoothing by convolution with a Gaussian distribution. A little further study is, however, necessary for two reasons. Firstly, the function $\bar{A}$ cannot be strictly Gaussian since it must fall to zero outside an interval set by the finite extent of the aerial and secondly, the approximating function is periodic.

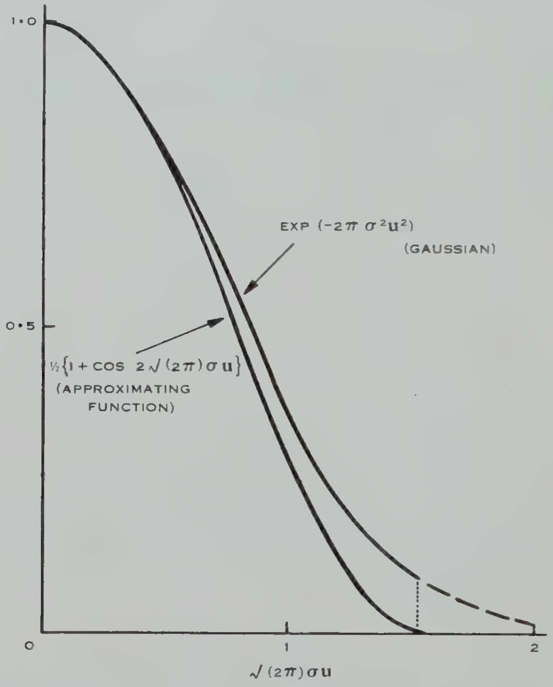


Fig. 4.—Agreement of the approximating function with the Gaussian distribution.

The periodic character of the approximating function

$$\tfrac{1}{2}\{1+\cos 2\sqrt{(2\pi)}\sigma u \cos 2\sqrt{(2\pi)}\sigma v\}$$

is obviously immaterial if $\bar{T}$ is zero for values of σu and σv greater than $(\frac{1}{8}\pi)^{\frac{1}{2}}$. Any method of ensuring that this is so would therefore be sufficient to render the periodic character inoffensive. One approach is to filter the data numerically, but a simpler procedure would often be sufficient. The values of T taken at critical intervals of u and v define a function T' whose transform is zero beyond the stated limits. If T' represents T adequately, within the limits of accuracy of the observations, then one may proceed immediately to operate on these values in the manner described in connexion with Figure 5.

The fact that the function $\bar{A}$ cannot be strictly Gaussian in practice leads to a step such as is shown by the dotted line in Figure 4. Possibly this is too small to be important in many cases, but circumstances will vary from case to case and it may be desirable at least to investigate the effect. An excellent fit to the stepped distribution can be obtained by adding a constant to the approximating function and altering the coefficient $\frac{1}{2}$ to compensate. The result will depend on the particular aerial considered, but for the 1500-ft Mills Cross it becomes, on the $v=0$ axis,

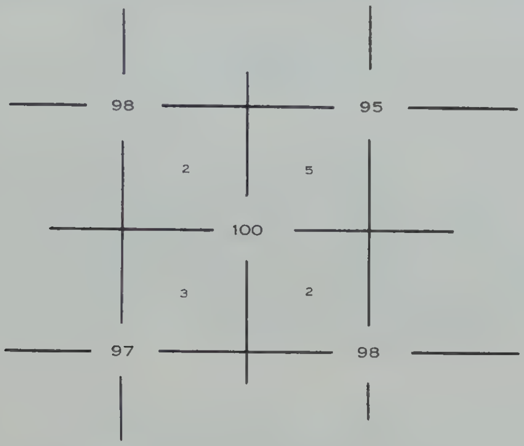


Fig. 5.—Illustrating the numerical procedure for smoothing.

$$0\cdot45\{1+\cos 2\sqrt{(2\pi)}\sigma u\}+0\cdot1.$$

The agreement of this function with $\exp(-2\pi\sigma^2u^2)$ is extremely close. The net effect is to weight the central value (100) when it is being averaged with the mean of the surrounding values.

VIII. The Calculation of Flux Density

As the result of a survey with a pencil beam, the distribution of observed brightness temperature T_a over a discrete source has been established and subtraction of the background is assumed to have been attended to. The flux density S of the source is given by integrating T over the solid angle occupied by the source, thus

$$S=\frac{2k}{\lambda^2}\int\int T\mathrm{d}\Omega.$$

We now show how to evaluate S from knowledge of T_a, first proving that

$$\int\int T\mathrm{d}\Omega=\int\int T_a\mathrm{d}\Omega,$$

The integral of a function between infinite limits is equal to the value of its Fourier transform at the origin, e.g. put $u=v=0$ in equation (4). Hence

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_a \mathrm{d}x\mathrm{d}y=[\bar{A}\,\bar{T}]_{u=v=0}$$
$$=[\bar{T}]_{u=v=0},$$

provided A is normalized so that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A\mathrm{d}x\mathrm{d}y=1$, when $[\bar{A}]_{u=v=0}=1$.

But

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T\mathrm{d}x\mathrm{d}y=[\bar{T}]_{u=v=0},$$

therefore

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_a\mathrm{d}x\mathrm{d}y.$$

It is consequently sufficient to calculate the flux density of a source by integrating T_a; nothing is lost through unawareness of the fine detail in T.

We now show that the integral can be replaced by a summation. If $\bar{T}_a=0$ for $|u|\geqslant\frac{1}{2}$ and for $|v|\geqslant\frac{1}{2}$, then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_a\mathrm{d}x\mathrm{d}y=4[{}^2\mathrm{III}(2u,\ 2v)*\bar{T}_a]_{u=v=0}$$
$$=[\overline{{}^2\mathrm{III}(\tfrac{1}{2}x,\ \tfrac{1}{2}y)T_a}]_{u=v=0}$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^2\mathrm{III}(\tfrac{1}{2}x,\ \tfrac{1}{2}y)T_a\mathrm{d}x\mathrm{d}y$$
$$=\sum_{\mu=-\infty}^{\infty}\ \sum_{\nu=-\infty}^{\infty} 4T_a(x-2\mu,\ y-2\nu).$$

Alternatively, *a fortiori*,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_a\mathrm{d}x\mathrm{d}y=\alpha\beta\Sigma\Sigma T_a(x-\alpha\mu,\ y-\beta\nu),$$

provided α and β are both less than 2. This result permits a great time saving: to integrate T_a, it suffices to sum only one in four of the discrete values necessary to specify T_a. The proviso that $\bar{T}_a=0$ for $|u|>\frac{1}{2}$ and for $|v|>\frac{1}{2}$ is automatically met when the units of x and y are equal to the peculiar intervals associated with $A(x, y)$, but if $\bar{T}_a(0, \pm\frac{1}{2})$ or $\bar{T}_a(\pm\frac{1}{2}, 0)$ are not all zero, then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_a\mathrm{d}x\mathrm{d}y=\frac{\Sigma\Sigma 4T_a(x-2\mu,\ y-2\nu)}{1+\frac{1}{4}\{\bar{T}_a(0,\ \frac{1}{2})+\bar{T}_a(0,\ -\frac{1}{2})+\bar{T}_a(-\frac{1}{2},\ 0)+\bar{T}_a(\frac{1}{2},\ 0)\}}.$$

If $\bar{T}_a$ is discontinuous at any of the points $u=0$, $v=\pm\frac{1}{2}$ or $u=\pm\frac{1}{2}$, $v=0$, the fact would be revealed as inequality between the four possible summations based on the points $(x, y)=(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ (their mean being correct). There are different ways of regarding this phenomenon. One can avoid it entirely by taking α and β less than 2. On the other hand one can use it for testing data supposed to be free of components of semi-period equal to the interval of tabulation.

IX. REFERENCES

BRACEWELL, R. N. (1955*a*).—*Aust. J. Phys.* **8**: 54–60.

BRACEWELL, R. N. (1955*b*).—*J. Opt. Soc. Amer.* **45**: 873–6.

BRACEWELL, R. N. (1955*c*).—*Aust. J. Phys.* **8**: 200–5.

BRACEWELL, R. N. (1956).—*Aust. J. Phys.* **9**: 198–217.

BRACEWELL, R. N., and ROBERTS, J. A. (1954).—*Aust. J. Phys.* **7**: 615–40.

SNEDDON, I. N. (1951).—"Fourier Transforms." (McGraw-Hill: New York.)

APPENDIX I

The Two-dimensional Fourier Transform of the Function ${}^2\mathrm{III}$

The integral $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^2\mathrm{III}(x, y)\mathrm{d}x\mathrm{d}y$ not being convergent, the function ${}^2\mathrm{III}(x, y)$ does not possess a regular two-dimensional Fourier transform: we are concerned here with showing that it has a transform-in-the-limit, and that this transform is identical with ${}^2\mathrm{III}$ itself. We prove in full the corresponding result for the one-dimensional function $\mathrm{III}(x)$, which has not been given before, to the author's knowledge, and indicate the method of proof for ${}^2\mathrm{III}$.

Consider the function†

$$F(\alpha, x)=\alpha^{-1} \exp\,(-\pi\alpha^2x^2) \sum_{n=-\infty}^{\infty} \exp\,\{-\pi\alpha^{-2}(x-n)^2\}.$$

For each α, $F(\alpha, x)$ represents a row of Gaussian spikes of width α, the whole multiplied by a Gaussian curve of width α^{-1}. We shall consider the limit of this function as $\alpha \rightarrow 0$. It is easy to show that $\lim F(\alpha, x)=0$ for $x \neq n$, and that, for every n, $\lim \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} F(\alpha, x)\mathrm{d}x=1$. It follows that‡

$$\lim_{\alpha\rightarrow 0} F(\alpha, x)=\mathrm{III}(x).$$

We now consider the transform of $F(\alpha, x)$, and show that as α tends to zero this transform approaches a limit which is $\mathrm{III}(s)$. In determining the transform

---

† In following this proof it will be helpful to notice that, in our $*$ and III notation, $F(\alpha, x)=\exp\,(-\pi\alpha^2x^2)\{\alpha^{-1} \exp\,(-\pi\alpha^{-2}x^2)*\mathrm{III}(x)\}$. Furthermore, $\exp(-\pi\alpha^2s^2)$ is the Fourier transform of $\alpha^{-1} \exp\,(-\pi\alpha^{-2}x^2)$; and the factors are so chosen that $\int_{-\infty}^{\infty} \alpha^{-1} \exp\,(-\pi\alpha^{-2}x^2)\mathrm{d}x=1$, irrespective of the value of α, thus ensuring that $\lim \alpha^{-1} \exp\,(-\pi\alpha^{-2}x^2)=\delta(x)$, whence, in due course, $\lim F\,(\alpha, x)=\mathrm{III}(x)$. The line of proof is to deduce, by conventional means, that $F(\alpha, x)$ has a regular transform, viz.

$$\alpha^{-1} \exp\,(-\pi\alpha^{-2}x^2)*\{\exp\,(-\pi\alpha^2s^2)\mathrm{III}(s)\},$$

and that both $F(\alpha, x)$ and its transform approach III as $\alpha\rightarrow 0$.

‡ This equation is to be interpreted in the same sense as one interprets $\lim \alpha^{-1}\Pi(x/\alpha)=\delta(x)$, namely, as permitting convenient notation as integrals for expressions which, when rigorously expressed, are limits of integrals.

of $F(\alpha, x)$ we first notice that the factor $\alpha^{-1}\Sigma \exp\{-\pi\alpha^{-2}(x-n)^2\}$ is periodic in x and may therefore be expressed as the Fourier series

$$\sum_{m=-\infty}^{\infty} \exp(-\pi\alpha^2 m^2) \cos 2\pi m x.$$

Hence

$$F(\alpha, x) = \sum_{m=-\infty}^{\infty} \exp(-\pi\alpha^2 m^2) \exp(-\pi\alpha^2 x^2) \cos 2\pi m x.$$

Applying the Fourier shift theorem we find for the transform of $F(\alpha, x)$

$$\bar{F}(\alpha, s) = \alpha^{-1} \sum_{m=-\infty}^{\infty} \exp(-\pi\alpha^2 m^2) \exp\{-\pi\alpha^{-2}(s-m)^2\}.$$

This function is a row of Gaussian spikes of width α with peaks lying on a Gaussian curve of width α^{-1}, and clearly,

$$\lim_{\alpha\to 0} \bar{F}(\alpha, s) = \text{III}(s).$$

Hence the transform-in-the-limit of III is identical with itself.

A similar proof can be given for ${}^2\text{III}$ by showing that the regular two-dimensional transform of $G(\Gamma * {}^2\text{III})$ is $\Gamma * (G\,{}^2\text{III})$, where

$$G(x, y) = \exp\{-2\pi^2\alpha^2(x^2+y^2)\},$$

and

$$\Gamma(x, y) = (2\pi\alpha^2)^{-1} \exp\{-(2\alpha^2)^{-1}(x^2+y^2)\}.$$

Here G and Γ are a transform pair, and Γ has the property $\lim \Gamma = {}^2\delta$. The doubly periodic function $\Gamma * {}^2\text{III}$ can be expressed as a double Fourier series and the two-dimensional shift theorem can be applied term by term. Both $G(\Gamma * {}^2\text{III})$ and $\Gamma * (G\,{}^2\text{III})$ approach ${}^2\text{III}$ as $\alpha \to 0$.

ON THE ASSOCIATION OF SOLAR RADIO EMISSION AND SOLAR PROMINENCES*

By J. P. WILD† and H. ZIRIN‡

[*Manuscript received April* 23, 1956]

Summary

Prominence cinematograms made at Sacramento Peak and Climax in the years 1949–55 have been examined and compared with the solar radio records at 167 Mc/s. No close connexion was found between limb events and radio events, but some eruptions were found to be associated with simultaneous radio bursts. Three such cases are discussed in detail.

The limb passages of large sunspots in 1949–55 were studied on both cinematograms and single exposure surveys with the view of finding a criterion for radio-noisy spots. It was found that spot groups showing looped prominences and downward streaming from the corona showed a marked tendency to produce radio noise storms. This result is ascribed to the fact that the seat of such storms must lie in the corona and in the presence of strong ordered magnetic fields.

I. INTRODUCTION

Our knowledge of the association of disturbed radiation from the Sun with activity observed in the optical spectrum is at present very limited. Considerable attention has been given to the relation between radio disturbances and transitory optical events observed on the Sun's disk ; but such events are seen mainly at the time of flares, whereas radio bursts frequently occur at other times. Coronagraph observations of the solar atmosphere reveal frequent activity in the form of sunspot prominences, eruptive prominences, surges, etc. at times other than flares. Most of this activity is seen in the corona at heights between about $0\cdot05R_0$ and $0\cdot5R_0$. This is just the region in which the radio emission is believed to originate.

In the present paper we summarize the results of a search for prominence activity associated with the generation of radio emission. Such an investigation is complicated by the fact that the limb is less favourable to the escape of radio waves compared with the central region of the disk. We have therefore approached the problem in two ways. Firstly, we looked for discrete transitory events on coronagraph records which coincided in time with bursts of radio emission. Secondly, we studied the general appearance on the limb of centres of activity which produced intense radio storms during their passage across the middle of the disk, and looked for persistent features which distinguished them from active centres which produced no storm. These two parts are discussed in Sections II and III respectively.

---

* This work was supported in part by the Air Force Cambridge Research Center, Geophysics Research Directorate, under contract AF 19(604)–969 with the High Altitude Observatory.

† Visiting High Altitude Observatory from Division of Radiophysics, C.S.I.R.O., Sydney.

‡ High Altitude Observatory of University of Colorado, Boulder, Colorado.

II. Eruptions on the Limb Associated with Bursts of Radio Emission

During an examination of prominence motion pictures recorded at Climax and Sacramento Peak in the years 1949–55, we noted three outstanding eruptions which coincided in time with isolated radio events and for which circumstances indicated that the optical-radio association was very likely to be real. These three events are described below.

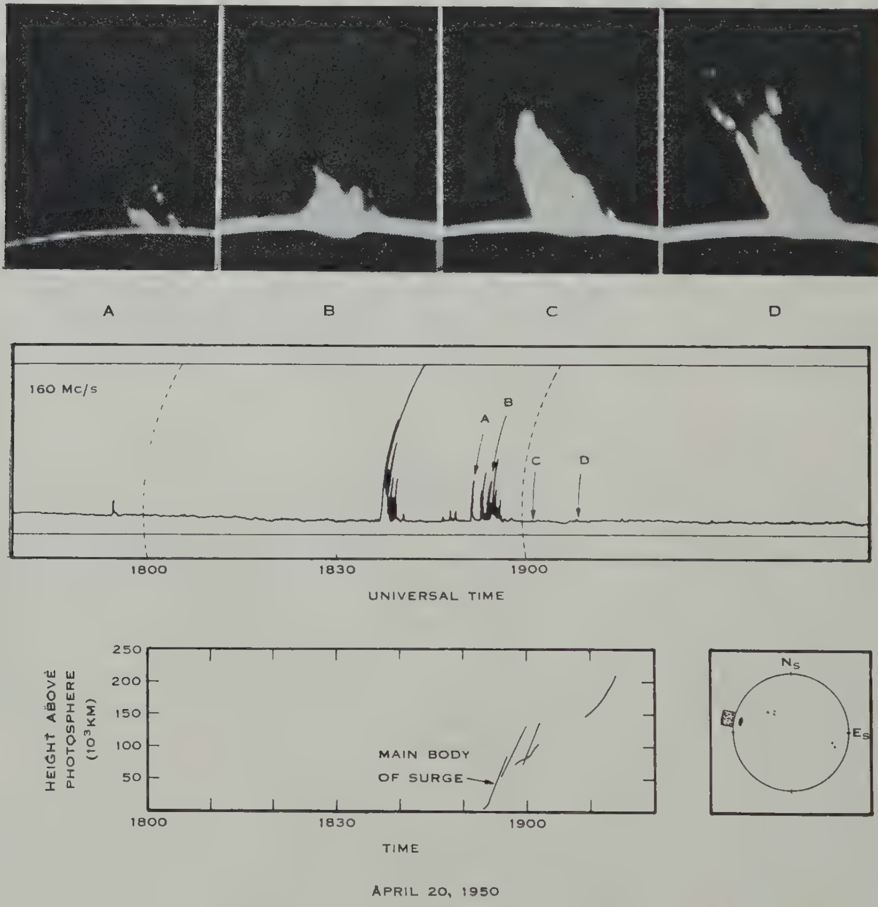


Fig. 1.—Great surge of April 20, 1950.

Great surge, April 20, 1950.—This was the largest surge noticed in our search. Details of this event are shown in Figure 1 which gives (i) a sequence of four coronagraph pictures, (ii) the radio flux record at 160 Mc/s, (iii) a plot of the projected height above the limb of certain clearly recognizable features of the surge, and (iv) a sunspot sketch. The Hα disk pictures taken at Boulder and the coronagraph pictures show the following sequence of events before and during the surge:

1805–1850 hr	Flare near the centre of the disk (maximum at 1820 hr)
1825–1835 hr	Sub-flare at active region near the west limb (see sunspot picture)
1852 hr on	Large surge originating in the vicinity of the large sunspot near the west limb, about 45,000 km (projected) inside the edge of the disk.

The main body of the surge rises with a projected velocity of about 300 km/sec, and fragments can be followed up to about 2×10^5 km before becoming invisible. The coronagraph observations were interrupted by clouds at about 1912 hr, and the return course of the surge, if any, is not known.

The radio event consists of a cluster of intense short-lived bursts at 1839 hr and a series of short-lived bursts starting at 1852 hr. From their appearance on the record, the first group is almost certainly of spectral type III, for which existing evidence points to an origin in very fast disturbances travelling upwards through the corona. The classification of the later series of bursts is doubtful, although their appearance suggests spectral type I (" storm " bursts). The level of the radio record remains practically undisturbed for at least $3\frac{1}{2}$ hr before and after this disturbance. Although it is difficult to fix the precise connexion of the bursts at 1838 hr with the optical events, the exact time coincidence of the series of bursts starting at 1852 hr with the first appearance of the great surge strongly suggests a causal relationship between the latter two.

Comparison of the radio record and height plot reveals an interesting conclusion: that the second part of the radio event takes place only during the early part of the surge, apparently while its leading edge is penetrating the lower levels of the corona. This is consistent with the notion that the visible matter follows immediately in the trail of the disturbance responsible for radio emission.

Surge-like eruption, May 2, 1950.—Details are shown in Figure 2. The limb event occurred near a large active area and appeared to constitute the eruption of a low-lying prominence located close to the sunspot prominence shown in Figure 4. The eruption initially resembled a somewhat nebulous surge (picture *B*) rising with a projected velocity of about 150 km/sec; from this matter, fragments subsequently rose to heights up to 200,000 km with velocities up to 250 km/sec. Such fragments are indicated by arrows in pictures *C* and *D*. An Hα disk picture showed no evidence of a flare at that time.

The radio level shows a gradual increase in activity immediately before the main eruption, increasing in intensity throughout the optical phenomenon and dying down to almost quiet level immediately after. In this respect the event differed from the preceding case; but it is noteworthy that fragments were being ejected throughout the main period of radio emission, so that again it is reasonable to associate the radio-producing disturbance with the leading edge of the visible matter. As before the coincident radio event appeared to consist of type I bursts.

Ascending prominence, January 18, 1955.—Details are shown in Figure 3. A quiescent prominence situated on the west limb at latitude 40 °N. was observed in the early stages of eruption when coronagraph observations began at 2030 hr (picture *A*). The top of the expanding arch rose with a mean velocity of more than 200 km/sec, and at 2050 hr (picture *C*) had reached a height of nearly 300,000 km. Observations ended at 2105 hr when downstreaming had begun (picture *D*). The eruption took place near the most active centre on the disk at that time, but no flare was reported from either Climax or Sacramento Peak.

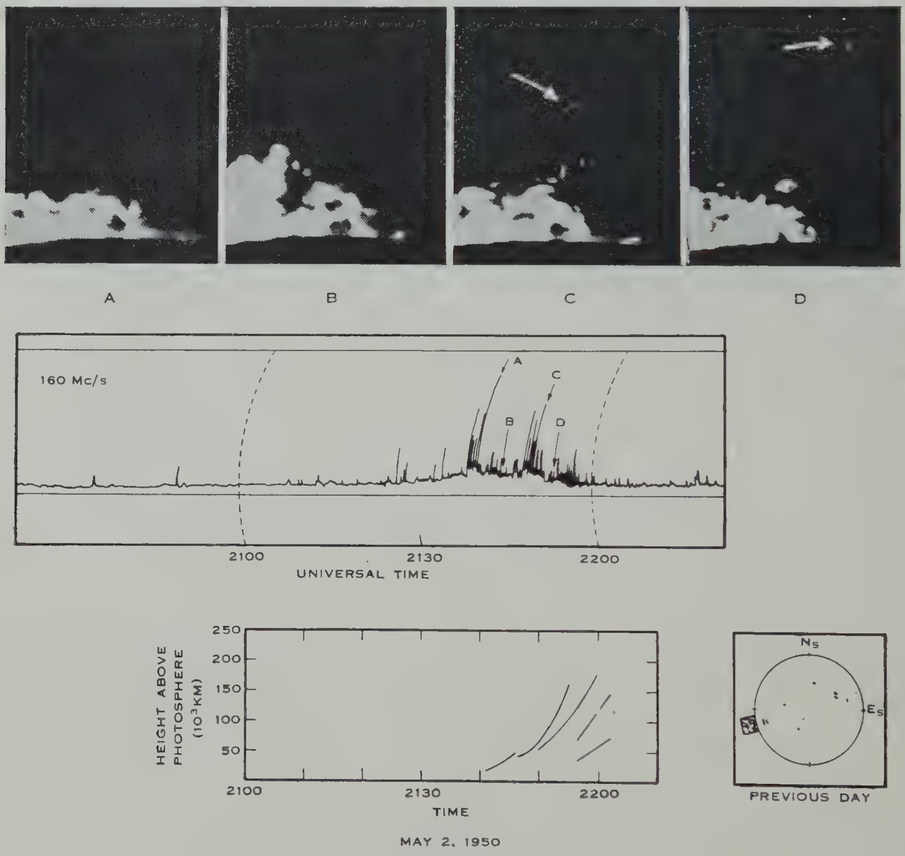


Fig. 2.—Eruption of May 2, 1950.

The radio record shows an extended storm, undoubtedly of spectral type I, beginning at 2000 hr and ending at about 2030 hr. Unfortunately the most intense radio emission occurs near 2130 hr after the coronagraph observations had ended. It is interesting that previous reports of ascending prominences by Dodson and Donselman (1951), Das and Sethumadhavan (1953), and Davies

(1953) also indicate that the most intense radio emission occurred in the late stages of the eruption.*

In addition to the three events described above, we noted numerous other eruptions at times for which 160 or 167 Mc/s records were available. The following is a summary of all cases noted :

	Surges	Ascending Prominences
Discrete radio events coincident in time . . .	8	2
Radio emission in progress, but association indefinite	23	1
No coincident radio event	27	1

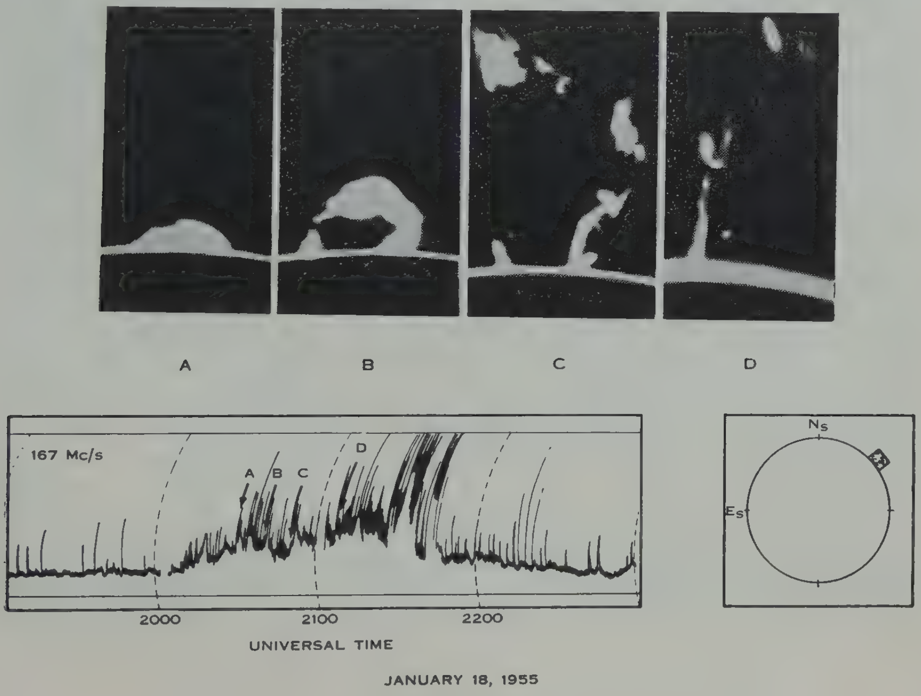


Fig. 3.—Ascending prominence of January 18, 1955.

It is seen that the proportion of surges which could reasonably be associated with discrete radio events is not high. This contrasts with the results of a survey by Warwick (1954) who found that 80 per cent. of surges observed coincided with unusual activity at 200 Mc/s.

In most of the cases for which the optical-radio coincidence was positive, the appearance of the radio event followed the general pattern of the three cases detailed above—i.e. suggesting a sequence of type I bursts.

---

* Dr. A. E. Covington has kindly informed us that a gradual rise and fall, starting at 2045 hr and lasting 45 min, was observed on 2800 Mc/s at Ottawa. The maximum flux was 10^{-21} W M^{-2} (c/s)$^{-1}$.

III. Prominence Activity Connected with Solar Radio Noise Storms

Radio noise storms are a phenomenon long associated with large sunspots. Payne-Scott and Little (1951), for instance, concluded that the presence of a sufficiently large sunspot in a spot group is necessary for the production of a storm. Furthermore, the observations of circular polarization in the storms and of the reversal of the sense of this polarization with change of hemisphere and of sunspot cycle is clear evidence of the importance of the spot magnetic field. The sunspot, however, is in the photosphere, and, as already noted, the radio emission originates from higher levels. We are therefore interested in the solar atmosphere directly above the sunspots.

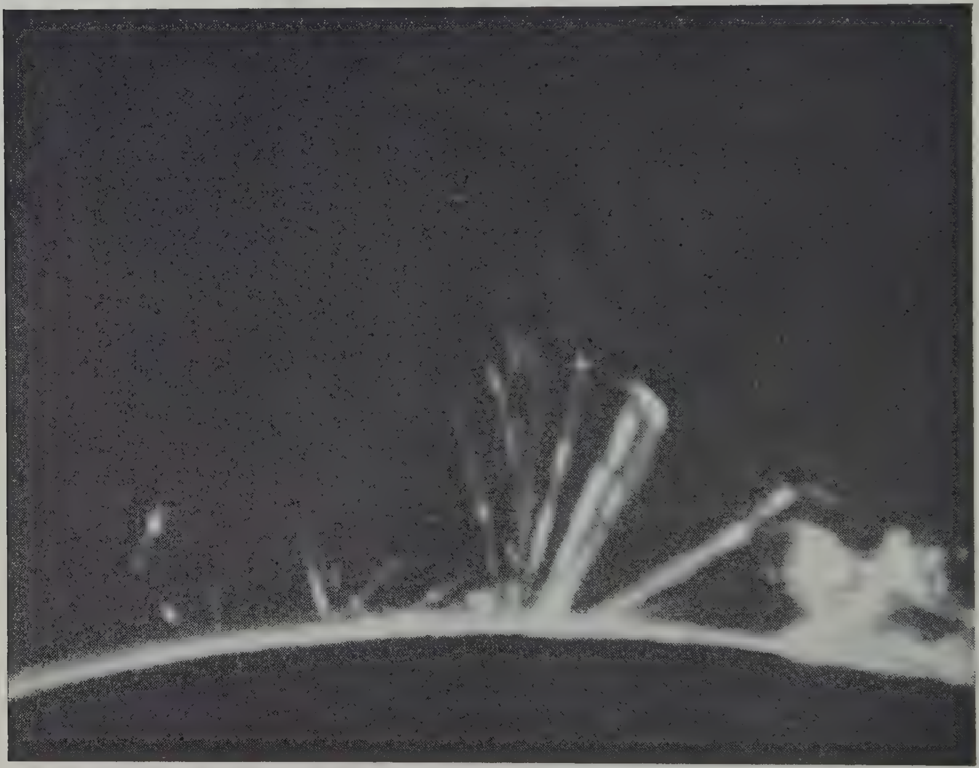


Fig. 4.—Typical complex loop prominence.

Payne-Scott and Little located by means of an interferometer the source of several large storms in 1949 and 1950 at certain spot groups. Examination of the limb appearance of these spots showed them to be accompanied predominantly by complex looped prominences and downward streaming from the corona. Interestingly enough, this sort of motion had not produced any radio coincidences at the limb, whereas the surges of the type discussed in Section II did not seem to be particularly characteristic of the radio-noisy spots.

We studied the prominence activity connected with each spot group in the following way. First, the CMP's of spot groups classified *E*, *F*, or *G* on the Zurich system were obtained from the daily medians of radio noise flux received from the Sun as reported in the *Quarterly Bulletin of Solar Activity.* Only those cases were retained where a clear decision was possible as to whether or not the spot group produced a radio storm. Then we obtained whatever Hα prominence films from Climax and Sacramento Peak were available, and also examined the daily Hα prominence surveys. In each case the appearance of the limb was classified with respect to the presence or absence of loop prominences and streaming from the corona. These correspond roughly to the classes AN_a and AS described by Menzel and Evans (1953). When those were present we measured the maximum height and estimated the importance. Streaming from quiescent prominences and ordinary quiescent prominences were considered negative cases on the basis of our original survey. Since the prominence motion pictures did not give very good coverage of all the dates (particularly for negative cases, since the observers customarily did not take prominence motion pictures when there was no interesting activity on the limb), the principal classification was done on the prominence surveys, which, between Sacramento Peak and Climax, gave quite good coverage. Although a motion picture is

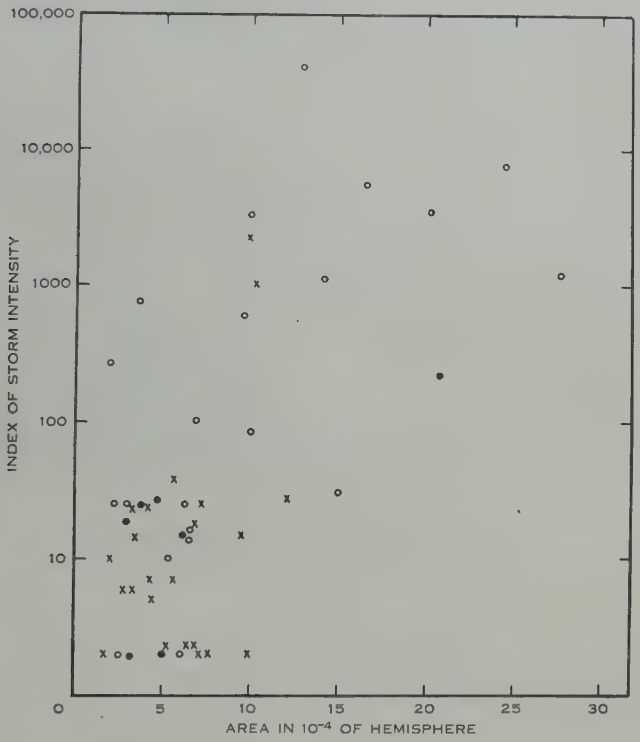


Fig. 5.—Radio storm intensity, sunspot group area, and limb characteristics.

○ Loops and/or streamers.
● E. and W. limbs contradictory.
× Quiescent, streaming from quiescent, or no prominences.

the best way to assay prominence activity, it is usually quite easy to interpret on the basis of experience a single survey photograph. Figure 4 is a good example of the kind of loops considered a positive case.

In view of the sharp contrast in magnitude in the radio record, a somewhat arbitrary index was used. The daily values in excess of 10^{-21} Wm^{-2} (c/s)$^{-1}$ flux at 62, 80, 98, 175, and 200 Mc/s respectively, were added for each day of the storm, being limited in all cases to within 5 days of CMP. Then the values on all five frequencies were added together to give an index of the intensity of the storm.

The scatter diagram of Figure 5 presents most of the data in graphical form. Storm intensity is plotted against spot group area, and each entry indicates whether the limb appearance was positive, negative, or contradictory (opposite classes from E. and W. limbs). The predominance of loops in the large storms is clearly marked. That this is not due to a correlation between spot area and loops is shown by the storms produced by small spots with loops. Since the prominences were observed 7 days from CMP, where the noise storm normally culminates, the correlation is surprisingly good. There is never any guarantee that a spot showing loops at one or both limbs would necessarily have them at CMP.

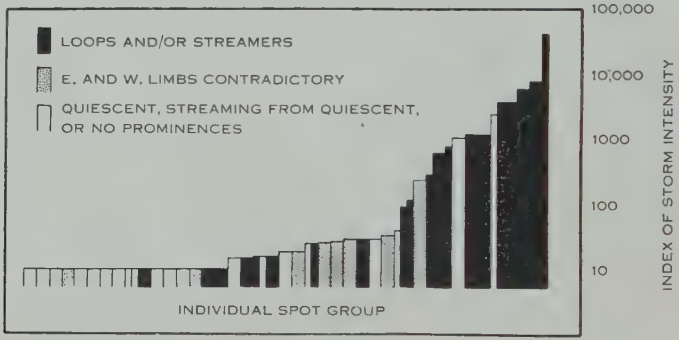


Fig. 6.—Spot groups arranged in order of increasing radio intensity, shaded according to limb characteristics.

The results of the investigation are presented in a somewhat different way in Figure 6. Figure 6 presents the various spot groups in order of increasing intensity and shaded according to limb appearance. Double width is given when observations were made at both limbs, and single when only one limb was observed. All spots with index of storm intensity less than 10 are grouped together, since distinction between low values is meaningless.

The general import of the data reported here seems to be that there is a close connexion between the ordering of the coronal material by the magnetic field, producing loops and streamers, and the production of radio noise storms. This agrees with the obvious implication of the observational facts at the beginning of this section, namely, that, in order to produce circular polarization, the magnetic field must extend up to the level of emission of radio noise. When we observe loops and streamers we know this to be the case.

IV. Conclusions

The results of this paper may be summarized as follows:

(1) While the connexion between visible activity on the Sun's limb and radio emission received at the Earth is not very close, certain eruptions cause almost simultaneous radio bursts at wave lengths of about 2 m. Such eruptions are characterized by the *outward* ejection of material, e.g. surges and ascending prominences.

(2) Analysis of two surges that generated radio emission suggested that the radiation originated from the *leading edge* of visible material—either the main body of the surge or ejected fragments—in its passage through the lower corona.

(3) An ascending prominence generated radio emission probably throughout the eruption; consistent with previous cases reported, the maximum radio intensity occurred late in the life of the eruption.

(4) While the active sunspot-prominences (characterized by downstreaming material from the corona and within high loops) produce no obvious increased radio emission when the active centre is on the limb, there is some evidence that they are responsible for radio storms during their passage across the central regions of the disk. The latter conclusion cannot be demonstrated conclusively owing to the limited and intermittent life of such coronal activity; but it has been found that the presence on the east limb of these active prominences provides a fair criterion for the occurrence of a radio storm around 7 days later; this criterion is at least as good as that based on the area of the spot group at the time of the storm. The hypothesis that the loop prominences are associated with the production of radio storms is attractive on the physical grounds that the seat of the storm must lie in the corona and in the presence of strong, ordered magnetic fields.

The prolonged radio storms, and possibly the transitory radio disturbances accompanying eruptions, consist of bursts. We therefore suggest that it is this type of radio emission which may be the most closely associated with visible disturbances in the corona.

V. Acknowledgments

We should like to acknowledge the use of the solar radio records of the National Bureau of Standards, through the courtesy of Mr. Robert E. Lawrence. Miss Dorothy E. Trotter, Mrs. Elizabeth Klein, and Mr. Robert Cooper of the High Altitude Observatory, were of great assistance in assembling and presenting the optical data. We are indebted to Dr. Walter Orr Roberts for his encouragement and support.

VI. References

Das, A. K., and Sethumadhavan, K. (1953).—*Nature* **172**: 446.

Davies, R. D. (1953).—*Nature* **172**: 446.

Dodson, H. W., and Donselman, R. W. (1951).—*Astrophys. J.* **113**: 519.

Menzel, D. H., and Evans, J. W. (1953).—Convegno Volta, p. 119.

Payne-Scott, Ruby, and Little, A. G. (1951).—*Aust. J. Sci. Res.* A **4**: 508.

Warwick, C. S. (1954).—*Astrophys. J.* **120**: 237.

IONOSPHERIC REFLECTIONS FROM HEIGHTS BELOW THE *E* REGION

By J. B. Gregory*

[*Manuscript received April* 6, 1956]

Summary

The pulse method of radio sounding has been used at 1·75 Mc/s to detect low ionospheric regions whose voltage reflection coefficient exceeds 4×10^{-6}. Observations from February to July 1955 have revealed reflections, from 95 km vertical height and of day-time occurrence only, which are believed to originate in the lower *E* region. A region with a lower boundary predominantly at 85 km has been found to exist continuously, and previous suggestions as to its meteor origin are queried. A series of complex partial reflections have been detected during day-time only down to 53 km, with significant heights near 75, 67, and 55 km. The occurrence and strength of these lower reflections increases markedly during winter.

From records of high absorption conditions during winter days, a connexion between the decrease in strength of *E* region reflections and the increase in strength of partial reflections between 66 and 88 km approximately is demonstrated.

I. INTRODUCTION

Experimental studies of the Earth's atmosphere have shown that radio waves are reflected not only from the *E* and *F* regions of the ionosphere, but also from heights below 100 km. Early observations of these low height reflections showed them to be weak and intermittent as compared with those from *E* and *F* regions. In recent years, apparatus of increased sensitivity has been used by several workers to gain more information about them. In this paper, an apparatus of high sensitivity and good resolution is briefly described, and the observations made with it on the lower ionosphere for a period of 6 months are presented.

II. APPARATUS

The apparatus makes use of the pulse method of radio sounding with transmissions on a fixed frequency of 1·75 Mc/s. A high sensitivity is obtained by means of a large aerial array, some 17 acres in extent; and good resolution results from the use of a short pulse of 8–10 μsec duration at half-power points. Records of radio reflections from various heights are obtained by photographing on continuously moving paper the intensity-modulated display of a cathode-ray tube. The apparatus is fully automatic in operation and is located at the Field Station of Canterbury University College Physics Department (lat. 43° 37′ S., long. 172° 24′ E.).

---

* Canterbury University College, Christchurch, New Zealand.

The main parameters of the equipment are:

Aerial System.—A vertically directed broadside array, with four rows of two collinear dipoles spaced λ/2 apart at a height of λ/7, is used. The beam-width between half-power points is 28° for the longer array dimension, and 46° for the shorter dimension. The calculated gain in the vertical direction over a half-wave dipole in free space is 12 dB. The array is used for reception as well as transmission.

Transmitted Pulse.—The peak power is from 3 to 5 kW. The pulse length is adjustable, the shortest pulse having a base of 25 μsec and a width between half-power points of 8–10 μsec.

Repetition Rate.—50 p.p.s.

Receiver.—The pass band is between 50 and 60 kc/s wide. Electrical noise at the site is usually low.

System Sensitivity.—A signal-to-noise ratio of unity is obtained for pulses returning from an infinite plane reflector, at 70 km height, of voltage reflection coefficient 4×10^{-6}.

Height Resolution.—Separation of partially reflecting regions less than 4 km apart is achieved.

Height Measurements.—Virtual heights are estimated to within ±1 km from the time delays of the leading edges of reflected pulses (see Section III (*b*)). Allowance for measured time delays within the transmitter and the receiver circuits is made by subtracting 5 km from all apparent heights.

III. Results

(*a*) *Nomenclature*

Although the term "*D* region" serves to identify that portion of the ionosphere below the *E* region, it is scarcely adequate for the description of the fine structure revealed in the present work. Accordingly, each reflection or series of reflections noted, whether due to permanent or temporary characteristics of the lower ionosphere, has been considered as arising from a separate region, which is distinguished from other such regions by stating the height at which its lower boundary is located. Future work may show that many partial reflections, particularly below 80 km, do not require to be thus separately distinguished; but a more formal designation may well wait until the mechanism of production of the reflections is better understood.

(*b*) *Height Interpretation*

The choice of the lower boundary as a means of identifying a region is influenced by the occurrence of many groups of partial reflections which show a spread in virtual height. Separate reflections within a group often show changes in virtual height of up to 10 km in the course of a fading period, so that individual heights may overlap. The position of the lower boundary of the group shows much less change and is usually estimated with sufficient accuracy by visual averaging over a length of the photographic records. The use of the leading edges rather than the peaks of reflected pulses for height measurements is consistent with this procedure.

For consistency, the heights of isolated reflections have been estimated at their lower boundaries. It may be noted, however, that some reflections are found to rise slowly and smoothly in amplitude, and, since the transmitted pulse resembles the Gaussian form, it is possible to decide the position of peak amplitude to within $\pm\frac{1}{2}$ km.

When the sensitivity of the system is decreased, for example, by the reduction of receiver gain, the lower boundary of a strongly reflecting region usually appears to move upward. This increase in height is more than can be attributed to the pulse shape or to change in receiver delay time, and, for the E region, may be as much as 10 km. For such conditions, the height of the lower boundary is taken as that at which the change of height with decreasing sensitivity first becomes small. For weaker reflections, since there is little apparent movement due to change in sensitivity, there is little uncertainty in the boundary position.

(*c*) *Estimation of Reflection Coefficients*

Preliminary measurements of the reflection coefficients at various heights have been made, and are given in succeeding paragraphs. Since multiple reflections from heights below the E region have not been found in this work, the ratio method of coefficient measurement could not be used. Further, the present aerial array has not permitted the separation of ordinary and extraordinary components of the reflected waves. The values given are based on measurements of voltage of the transmitted and the received signals, and on a calculation of aerial gain. Future measurements may require some revision of these results, but it is likely that the relations between values for different heights will be little changed.

For automatic recording, measurements of the strength of reflected pulses are made by reducing the receiver gain in a series of 12 steps, each of about 6 dB, throughout a 6 min period once every 30 min. The effect of this reduction on the photographic records may be seen in Plates 1 and 2. Calibration of these steps has been made from time to time, and the data obtained photographically have been checked and supplemented by amplitude measurements on a monitor cathode-ray tube.

(*d*) *Extent of Observations*

The results given are based on observations made on 130 days between January 27 and July 31, 1955. They therefore include summer, equinox, and winter conditions. A daily transmission covering the period 1030–1330 hr L.M.T. is made ; and this has been supplemented by some 30 transmissions of up to 3 days' continuous operation. Night transmissions are included in this latter group. Some gaps exist in the records due to development work in the concluding stages of construction, and to minor breakdowns thereafter. The use of automatic stepped gain control of the receiver, which also allows the recognition of separate regions in apparently continuous reflections, commenced on May 17. Prior to this date, night observations were less frequent.

(*e*) *Main Reflecting Regions*

(i) E *Region.*—The strongest reflections recorded have been from the *E* region, at heights of 105 km and above. Near 95 km, there have appeared reflections which vary between partial and total, and which in the latter condition are as strong as those noted at 105 km at other times. In subsequent paragraphs, reasons are given for considering these 95 km reflections to originate in the lower part of the *E* region. During daylight hours, reflections from above 95 km have invariably been present, even though, under abnormal conditions, they have been as much as 15 dB weaker than reflections from lower heights. Thus all the reflections below 95 km noted in this investigation have been partial and not total.

Examination of the records has revealed several recognized phenomena of the *E* region, such as boundary height variations; apparent movements of reflecting centres with velocities of the order of 100 m/sec; rapid fading of reflections at greatest range; and the occurrence of sporadic *E*. Such effects have not been studied in detail, but two other phenomena, namely, meteor reflections and the reflections at 95 km height, are worth description.

Table 1

95 KM REGION SUMMARY

Lower boundary height :	94–97 km, predominantly 95 km
Typical reflection coefficients	(1) For total reflection, 5×10^{-3} (2) For partial reflection, 5×10^{-4}
Occurrence :	Day-time only, on single or associated days in irregular sequence
Presence :	Either (1) Continuous, for all daylight hours or (2) Intermittent during day-time

(1) *Meteor reflections.*—Plate 1 (*A*) shows examples of " bursts " of reflections at heights above 100 km. These bursts last for periods of from a few seconds to one or two minutes, and are more easily visible in the absence of *E* reflections. The majority of them show an abrupt rise in amplitude followed by a roughly linear decrease in width of the recorded trace. The resemblance between these bursts and those detected by Appleton and Naismith (1947) during the Giacobinid meteor shower is very marked. The pulse lengths used in each experiment are comparable; but the meteor transmissions were made at 27 Mc/s. Despite the difference in frequencies, it is clear that the reflections shown in Plate 1 (*A*) are also of meteor origin. The lower limit to the heights at which they appear is of interest in connexion with a region described in Section III (*e*) (ii) and investigation has shown that they are not detected below 100 km. Measurement of their reflection coefficient by the present method requires coincidences between bursts and the various steps of receiver gain reduction. The highest value noted so far has been about 10^{-4}, but this may well be exceeded.

(2) *The* 95 *km region.*—An example of total reflection at 95 km height appears in Plate 1 (*B*), where a region extends beyond the ninth step of gain reduction, corresponding to a reflection coefficient of about 5×10^{-3}. A summary of the main features of this region is given in Table 1.

The 95 km region appears to be a part of the *E* region, in contrast to other regions at still lower heights. Not only does it give reflections which are as strong on occasions as those to be expected from the *E* region at the same time of day, but its formation from the *E* region by a downward splitting has been observed. For example, on March 14, 1955, at noon, the lower region separated out of the *E* region at 103 km until, in 9 min, its boundary had reached 94 km. At 94 km, the region slowly weakened and vanished ; and at 1450 hr, the whole process was repeated. Alternations of the effective boundary of the *E* region between a higher value and 95 km occur frequently, with total reflection from the lower height. On many days, 95 km reflections are absent, and a slight gap appears in the record, indicating that any residual reflections from this height are some 60 dB weaker than the strength at which the region gives total reflection. Again, the 95 km region is often present as a partial reflector, and has been identified at a strength of 30 dB less than that of the *E* region. These observations are consistent with the formation, below the height of maximum ionization density in the *E* region, of a " ledge " or secondary maximum of ionization whose density may sometimes exceed the critical value of $3 \cdot 77 \times 10^4$ electrons/cm^3.

Reflections from 95 km approximately were noted on 44 days of May, June, and July. The presence of the region may be inferred from records obtained in the preceding 3 months ; but, as automatic gain reduction was not then in use, the heights of the lower boundary are not available. The periods of transmission throughout the 6 months were not distributed in a sufficiently regular manner to warrant an analysis of the diurnal variation of the time of occurrence of the region, but some trends are evident. Days during which the region is continuously present as either a partial or a total reflector from sunrise to sunset are much less frequent than those in which it appears for an hour or two only. These latter appearances may be at any daylight hour, but there is a tendency for the region to appear for several days in succession between 12 and 1330 hr L.M.T., and to persist on such occasions for not more than one hour.

Finally, it is desired to point out that the foregoing applies only to conditions under which there is a recognizable lower boundary of this region. There exist also at other times conditions of diffuse reflections which are continuous from lower heights through 95 km to the *E* region, but these have not been included in the discussion.

(ii) *The* 85 *km Region.*—The most persistent region detected in this investigation lies below the *E* region, with a lower boundary close to 85 km. Examples of records in which this region may be seen are shown in Plate 2. A summary of its main features is given in Table 2.

The 85 km region has been noted during both day and night in summer, autumn, and winter, and appears to have an almost continuous existence. It has been identified most clearly in records which include the use of automatic gain reduction. This method was introduced in May ; and it is therefore convenient to discuss the behaviour of the region during February, March, and April separately from that during the succeeding months.

In the first 3 months, more than 30 day-time and several night-time transmissions of varying durations were made. The region was present throughout each of the former, and for the greater part of each of the latter. It may have been absent before sunrise on several summer nights, when it could not be detected in the presence of interference from distant broadcasting stations. There is also a possibility that the failure to detect the region at these times was due to equipment faults, and further observations will be needed to settle this point. The available data are, however, more than sufficient to establish the existence of the region at all hours of day and most hours of night throughout these months.

TABLE 2

85 KM REGION SUMMARY

Lower boundary height	: Predominantly 85 km
Occurrence	: Day and night (summer, autumn, and winter observations only)
Presence	: Continuous (see text)
Typical reflection coefficients	: (1) Noon, 4×10^{-4} (2) Night, 8×10^{-5} to 8×10^{-4}
Diurnal height variations	: (1) Summer; within ±3 km during the day, with a sunrise decrease. Night data are not available (2) Winter; a possible noon minimum, and within ±3 km during the night (see text)
Seasonal height variation	: A possible decrease of 2 km in winter

The day-time height of the lower boundary was usually between 85 and 88 km. Changes in height within these limits have been found to include a rise or a fall throughout the daylight hours, or a tendency to a minimum at noon.

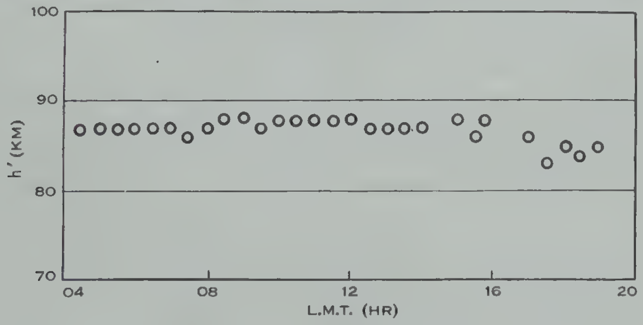


Fig. 1.—The lower boundary height of the 85 km region on March 18, 1955.

Data on summer night conditions are insufficient to establish the nature of height changes after sunset. The behaviour of the region during sunrise periods on different days showed some variation. On 6 days, the region persisted throughout the sunrise periods without discernible change in height. Figure 1 illustrates this type of behaviour. On 15 days, the region decreased in height after sunrise by an amount which varied between 1 and 12 km on different days. Each decrease was gradual, and occurred throughout a period of about

1 hr after ground sunrise. Figure 2 shows the lower boundary height for one such day. It is possible to suggest two extremes of behaviour with which this decrease may be associated. In the first, the region is assumed to persist with only slight reduction of reflection coefficient throughout the night, and to rise slightly prior to sunrise. In the second, the region is assumed to vanish before sunrise, to re-form at about 95 km, and to strengthen at successively lower heights until, after a lapse of an hour, it reaches the day-time height. Records have been obtained which support each of these possibilities and also intermediate behaviour. In particular, the largest values of decrease, which occurred in March, seem to be associated with the re-formation of the region. A lack of reflection coefficient values during February, March, and April has prevented a more detailed study of this effect.

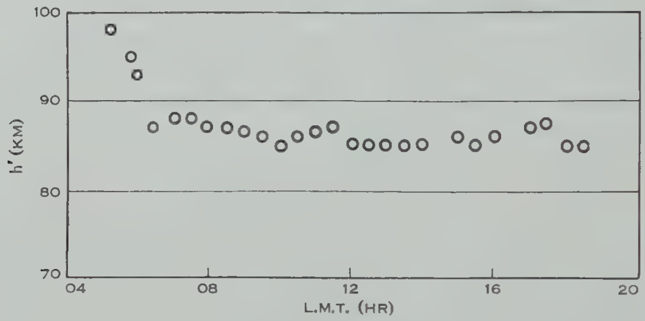


Fig. 2.—The lower boundary height of the 85 km region on March 22, 1955.

During the months of May, June, and July the day conditions differed markedly from those of earlier months. This difference is not confined to the 85 km region alone, for during winter there is a considerable increase in the strength of reflections from below 85 km and also a decrease in the heights from which these reflections are received. It is scarcely possible to consider the 85 km region apart from these lower reflections, which are discussed in the following section, but its condition usually falls into one of three categories. The first of these is as shown in Plate 2 (*B*). Despite the presence of reflections from immediately below it, the boundary of the region is easily recognizable from the increase in reflection strength. On 49 days of May, June, and July its height was found to be between 80 and 88 km, with the majority of values at 85 km. The second category is represented by a less abrupt change in reflection strength, so that the existence of a separate boundary is open to question. An extreme example of this condition, in which continuous diffuse reflections cover the heights in question, is shown in Plate 1 (*B*). Some 28 days throughout the 3 months may be included in this category. The third category is of considerable interest, as it is found to include the well-known winter days of high absorption. It is discussed separately in Section III (*f*).

A comparison of record (*A*) of Plate 2 with record (*B*) of the same plate shows that the lower boundary of the 85 km region was less definite on winter days than on summer days. The data quoted in the preceding paragraph show, however, that the region retained its identity on a majority of days. On days when the position of the boundary could not be located, reflections were still obtained from above 85 km, and at sunset, the region re-emerged. It seems safe to conclude, despite the limited periods of transmission, that throughout the winter the region had a continuous existence. Further, when the summer evidence is considered, it seems likely that the region will be found almost continuously present throughout the year. Hence it may be inferred that the processes which cause it differ from those causing the winter increase of reflections at lower height.

The presence of these latter reflections complicates an attempt to determine the height variations of the 85 km region, and a study of the ionization density in the lower ionosphere is to be preferred. In this connexion, the records of a number of winter days show that reflections of the same strength as those received from the 85 km region in the early morning are also received from heights which, starting at 85 km, decrease steadily throughout the morning to a minimum at noon and then rise again throughout the afternoon. This movement may be regarded as a real change in the 85 km region boundary; but it may be considered as an apparent change only, which is due to the formation of extra ionization below a constant 85 km region. The distinction is probably artificial, but the second approach serves to emphasize the possibility of two different processes operating on a limited region of the lower ionosphere. Further reference to extreme cases of this decrease in height is made in Section III (*f*) (i).

Reflections from below 85 km approximately have been received only in day-time, and hence there has been little difficulty in detecting the 85 km region during winter nights. It was continuously present throughout four full night transmissions, and was never absent from many shorter night transmissions. The lower boundary was usually between 81 and 85 km, and often remained stable to within ±1 km or less for periods in excess of an hour. Changes in height tended to be from one discrete height to another, rather than through a series of continuous heights. The region either maintained a nearly constant strength throughout the night or weakened in the latter portion of the night to a minimum strength before sunrise.

The seasonal variation of height of the 85 km boundary may be inferred from Figure 3, which shows a plot of lower boundary heights for the 6 months February-July. In this figure vertical lines indicate the heights throughout which continuous reflections, without sharp boundaries, were present. Although the data for summer months are incomplete, it is possible to infer a drop of average boundary height by some 2–3 km in winter.

(iii) *Regions below* 80 *km*.—Reflections have been detected from below the lower limit reached by the boundary of the 85 km region (about 80 km) down to a minimum height of 53 km. It is usual to record at least one, and sometimes three separate heights of reflection within this range at any time of day. Wide variations exist in the heights of the boundaries, the durations, the mean strengths, and the fading rates, so that the summary (Table 3) is less satisfactory than those for higher regions.

These lower reflections are visible in Plates 1 (*B*), 1 (*C*), 2 (*A*), 2 (*B*), although a much wider variety of record pattern is encountered. The outstanding feature of this portion of the ionosphere is the pronounced increase in the vertical extent and the strength of reflections during winter days. In this period, there is present, from about 55 km upwards, a series of regions whose separate identity is readily distinguished on the majority of days, but which sometimes coalesce into one continuous region. Figure 4, which is a histogram of lower boundary heights at half-hourly intervals, indicates the distribution of the regions throughout 6 months. (In order to maintain a reasonable comparison between successive months, the data used in this histogram have been limited to readings taken between 1030 and 1330 hr L.M.T. Data for this period during February are inadequate for proper comparison, and all available readings have been used.)

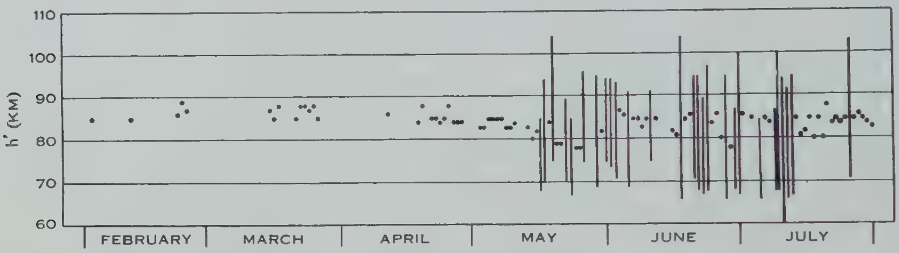


Fig. 3.—The noon heights of the 85 km region lower boundary for the months February-July 1955.

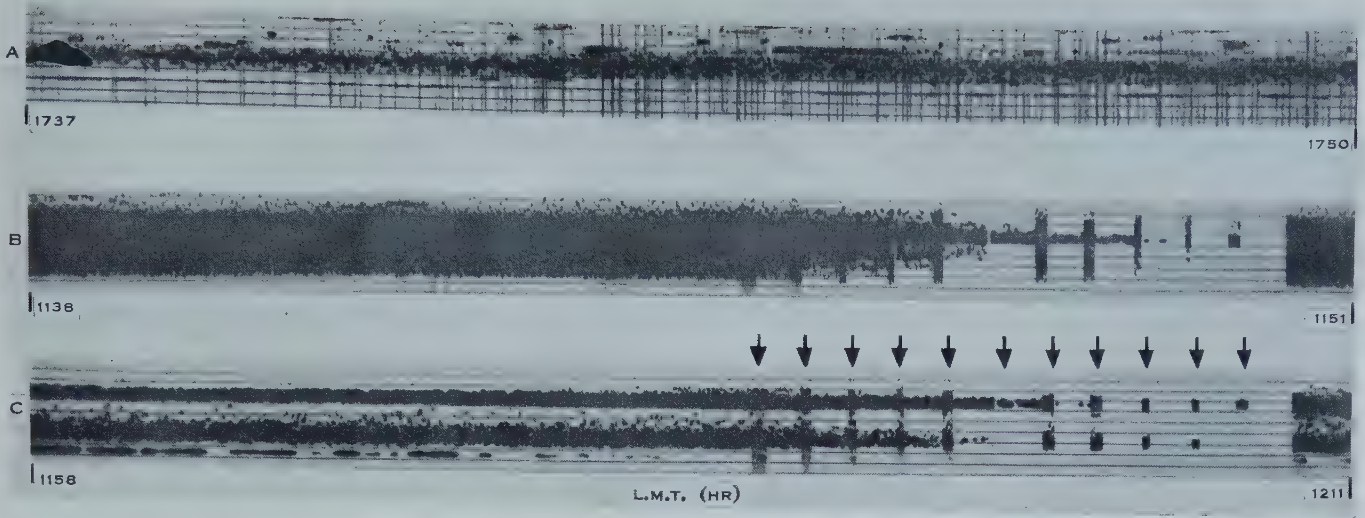
TABLE 3

SUMMARY OF REFLECTIONS BELOW 80 KM

Significant heights (km) :	(See Fig. 4)
February	75
March	75, 70
April	75, 70, 65
May	75, 70, 67, 65, 55
June	75, 70, 68, 57
July	—, —, 68, 57, 56
Reflection coefficients :	Decreasing with decreasing height to a minimum of 5×10^{-6}
Occurrence :	Day-time only
Presence :	Intermittent
Height variation of lower boundaries :	Lowest reflections, irregular within ±2 km Higher reflections, slight noon decrease

On a typical winter day, the reflections down to about 65 km form one group, with a reflection coefficient of about 4×10^{-5} at that height. Below this height, there is usually a reflection-free gap of up to 10 km, and then a final region at about 56 km. The coefficient of the latter rarely exceeds 2×10^{-5},

IONOSPHERIC REFLECTIONS BELOW THE E REGION



h′t Records

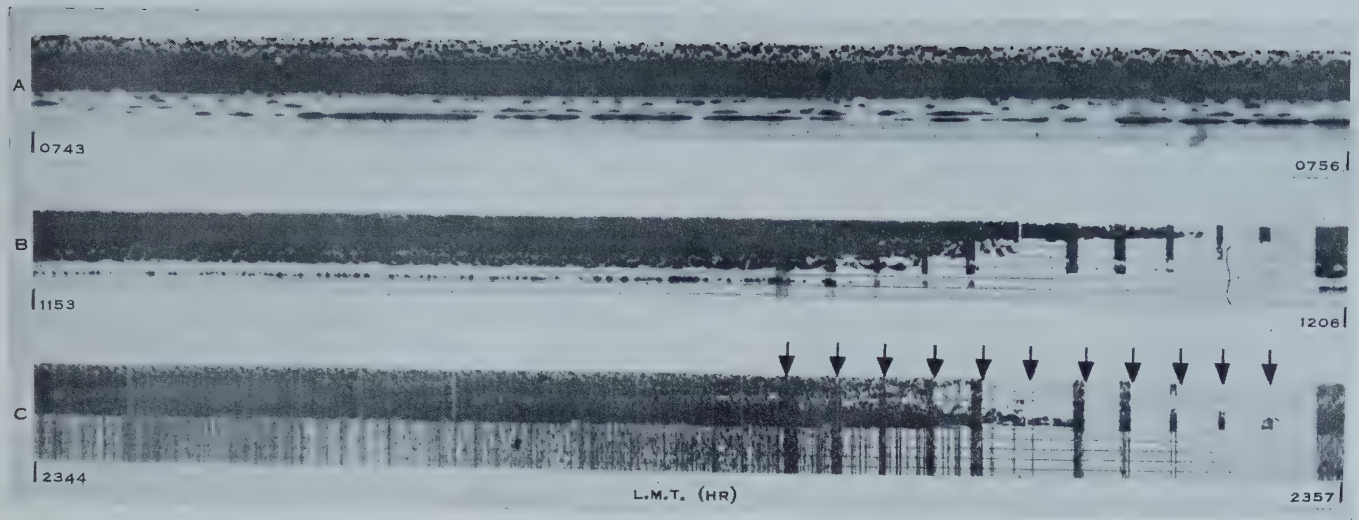
(*A*) March 26, 1955. Penetration of the *E* region, showing meteor reflections above 100 km. The 85 km region and fading reflections from 74 km are visible.

(*B*) July 11, 1955. Total reflection occurs at 95 km ; and continuous reflections, down to 59 km, are present.

(*C*) July 23, 1955. High absorption conditions ; the weakened *E* region reflections, the " gap " above 85 km, and strong partial reflections at 67 km and above are all visible.

(Note : The first marker is at 45 km corrected height, and the remainder are at 10 km intervals. An arrow marks the start of each step of gain reduction, which is preceded by a short switching interval at higher gain.)

IONOSPHERIC REFLECTIONS BELOW THE E REGION



h′t Records

(*A*) February 23, 1955. Continuous reflections from 88 km and fading reflections at 74 km are visible. At the left of the record a 67 km sporadic reflection, of 18 min duration, commences.

(*B*) July 19, 1955. Typical winter conditions, including the *E* region, the 85 km region, and lower reflections, are shown.

(*C*) July 18, 1955. The stable boundary of the 85 km region at night is visible through broadcast interference.

(Note : The first marker in (*A*) is at 55 km corrected height, and in (*B*) and (*C*) it is at 45 km.)

and is often barely detectable. From considerations of pulse length and the apparent thickness on the records, the thickness of the reflecting portion of this lowest region cannot exceed 1 or 2 km. The fading rate at any height, though subject to wide variation, is generally rapid and may have a period of a few seconds only. In particular, the lowest region is subject to large day-to-day variations in fading rate.

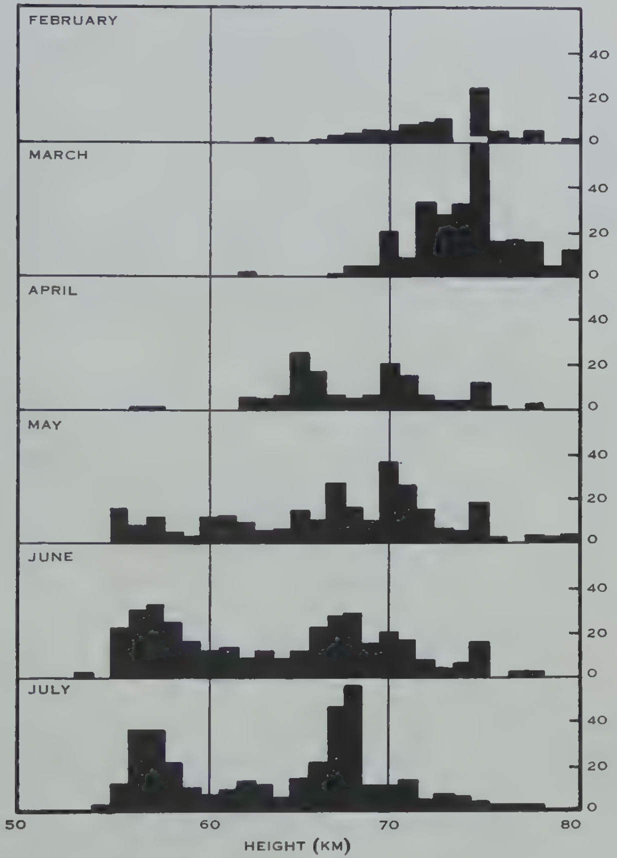


Fig. 4.—Histogram of occurrences of reflections from below 80 km height between 1030 and 1330 hr L.M.T. each day, using readings taken at half-hourly intervals.

In summer, sustained reflections are much less frequent below 70 km. They are weaker than those from similar heights during winter and appear to arise from sharp discontinuities. They are quasi-periodic and for a high proportion of the time are below recording strength. An interesting feature of summer conditions is the sporadic appearance of isolated reflections of short duration. At some particular height such a reflection may appear once for 15 sec during the daylight period, rising to a single maximum and then fading completely. Longer durations are common and usually include several maxima with a period of the order of a minute. Plate 2 (*A*) shows the start of one such reflection, of total duration 18 min. Not more than three per day have been recorded, and data are insufficient to attempt an analysis. An estimate of their reflection coefficient sets a figure of 7×10^{-5} as a maximum value.

(f) *High Absorption Conditions*

(i) *Records showing High Absorption.*—On 12 days it has been possible to note a marked decrease in the noon reflection coefficient of the E region. Photographic records of this decrease are supported by visual observations made on monitor equipment 22 km from the transmitter. A comparison of the strength of obliquely reflected waves with the unchanged strength of a ground wave showed that the decrease was real. The weakening of E reflections is attributed to high absorption and is usually accompanied by a strengthening of partial reflections below 85 km. Continuous observations throughout some of these days have shown that the 85 km region is present at sunrise, sometimes at a lower height than usual, and that there is an apparent decrease in height of the boundary of the 85 km region to as low as 66 km at noon, with a gradual rise thereafter until the original height is regained about sunset. Figure 5, which is a plot of boundary height during May 17, illustrates this behaviour. Above the heights shown in Figure 5 there exists a series of continuous partial reflections, whose vertical extent has been found, after allowance for pulse length, to vary between 12 and 20 km on different days. Within this series of reflections the reflection coefficient shows an increase over that at similar heights on days of lower absorption, with greatest strengthening from about 70 km upwards. Typical values range from 3×10^{-5} at 67 km to 4×10^{-4} at 75 km and above. On high absorption days the records show an apparent upper boundary to the series of reflections at a height which lies between 85 and 95 km. Between this boundary and the E region there appears a characteristic " gap " in the record, in which there appear only brief isolated reflections whose noon coefficient does not exceed 3×10^{-5}. An example of high absorption conditions is shown in Plate 1 (C).

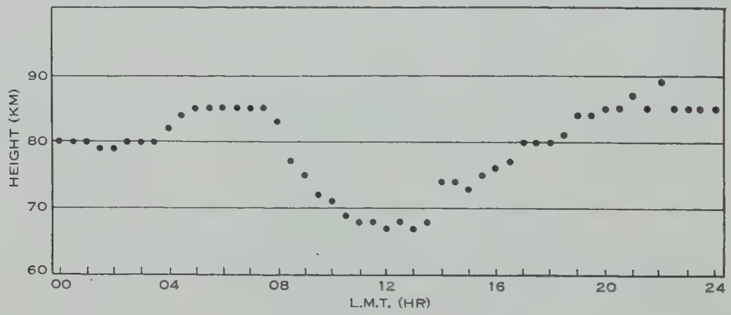


Fig. 5.—Plot of the lower boundary height of the 85 km region throughout a high absorption day, May 17, 1955.

(ii) *The Location of the Highly Absorbing Region.*—A qualitative relation between the enhanced partial reflections and the increased absorption described in the previous section may be deduced in the following manner. At any height, the partial reflection coefficient, ρ_p, is proportional to the change, ΔN, in electron density, N. Magneto-ionic theory predicts that the absorption coefficient, K, is proportional to N. For the heights in question, Gardner and Pawsey (1953) have concluded that, under low absorption conditions and possibly more generally, ρ_p is proportional to N.

Thus

$$\rho_p \propto K,$$

and further,

$$\int \rho_p.\mathrm{d}h' \propto \int K.\mathrm{d}s,$$

where the latter quantity is the total absorption for the given range of heights and the former is the integrated partial reflection coefficient over the same heights. Hence, if the computed values of $\int \rho_p.\mathrm{d}h'$ for a given range of heights are found to be proportional to the experimental values of $\int K.\mathrm{d}s$, the inference is that a constant fraction, if not all of the absorption is located within these heights.

The information available from records such as shown in Plate 1 (*C*) is sufficiently detailed to warrant a search for a similar proportionality under high absorption conditions. Values of ρ_p at different heights have therefore been obtained for the noon period of each of seven high absorption days. After allowance for the duration of the transmitted pulse, the heights showing enhanced reflection were found to be within the limits of 66 and 88 km. These were subdivided into increments of constant ρ_p and the mean value of $\Sigma\rho_p.\Delta h'$ was computed for the noon period of each day from three sets of gain reduction. In the same way, the mean strength of E region reflections was found, and its value, expressed in decibels, was taken as a measure of $\int K.\mathrm{d}s$. Separate values of $\Sigma\rho_p.\Delta h'$ on any one day were never further apart than 50 per cent. of the mean, and were often less than 10 per cent. apart. E region strengths were estimated to within ± 2 dB.

A plot of the two quantities is shown in Figure 6, and a connexion is evident. This figure is supplemented by Table 4, which gives details of the heights of the regions concerned.

The relation established in Figure 6 for heights between 66 and 88 km does not preclude a proportionate contribution to the total absorption from other regions, including the E region. The existence of other contributions to total absorption is evident from the fact that, on several winter days, the average strength of E region reflections during the midday period has exceeded 60 dB; whereas the value obtained by extrapolating Figure 6 to zero partial reflection coefficient below the E region is only 55 dB. Below 66 km, the amount of absorption is believed to be small, since only very weak reflections have been detected at these heights at any time. The conditions above 88 km and below the E region on high absorption days are discussed in the next section; but information on the extent of absorption above 88 km was not available. The appearance of records such as shown in Plate 1 (*C*) does suggest that the major portion of the high absorption is located at heights such as shown in Table 4, and whose limits appear to be within 66 and 88 km.

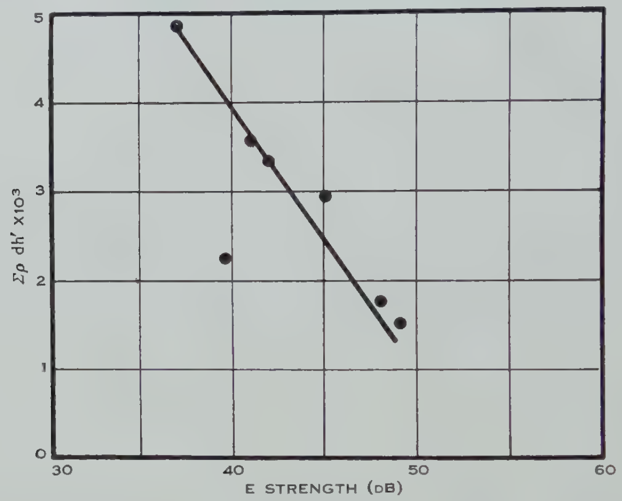


Fig. 6.—Plot of the integrated partial reflection coefficient, between 66 and 88 km limits, versus the strength of the *E* region at noon on separate high absorption days.

(iii) *The Absence of Reflections between* 85 *km and the* E *Region*.—A second aspect of the high absorption days concerns the " gap " between 85 km approximately and the *E* region. It is of interest to enquire whether this is due to a real absence of ionization or whether reflections from ionization actually present are weakened by absorption in the region below. If established, the first possibility would support an interpretation of Figure 5 in terms of a mass downward movement of ionization by some 20 km into regions of increased

Table 4

HIGHLY ABSORBING REGION

Date	Absorbing Region		$\Sigma\rho_p.\Delta h'$ $\times 10^5$	*E* Region Strength (dB)
	Lower Boundary (km)	Vertical Extent (km)		
17. v.55	67	15	175	48
23. v.55	70	17	488	37
21. vi.55	71	17	294	45
22. vi.55	68	20	335	42
23. vi.55	67	16	356	41
6.vii.55	66	18	223	39
9.vii.55	67	19	150	49

collision frequency. The establishment of the second possibility would require the explanation of Figure 5 in terms of an additional ionizing agency.

Alternative approaches have been made to this problem. The first assumes that both the *E* region and the 85 km region conditions remain as on normal days ; but, that due to high absorption below them, reflections from each weaken equally. Estimates of the relative decreases have given inconclusive results, and the lack of an independent check on *E* region conditions has prevented the method from being carried further. The second approach takes note of the fact that random reflections occur right up to the *E* region boundary. These reflections must originate either at vertical incidence and thus at a true height very close to that recorded, or at oblique incidence from irregularities within the region below 85 km. In several records, reflections may be found which, if oblique, would require a path whose angle with the vertical is not less than 45°. For such a path, the directive characteristics of the aerial array would cause a reduction of reflection strength by at least 28 dB in comparison with that for vertical incidence. The reduction of the observed reflections is never more than 22 dB below the maximum strength for a normal 85 km region, and is sometimes only 15 dB less. The predicted strength for an oblique path is thus considerably exceeded and it appears that the origins of the random reflections are at heights between 85 and 105 km. The short duration of these reflections is attributed to the rapid fading often noted at such heights, so that only the maximum strength exceeds the value which can be detected through the absorption below. The continued existence of an 85 km region, as described in Section III (*e*) (ii), throughout high absorption days is thus supported, and the apparent decrease in its boundary height as shown in Figure 5 may be attributed to an increased ionization density in regions below it, due to an additional ionizing agency.

(*g*) *Day-to-day Variations*

The overall appearance of the photographic record for any particular day is determined by the number of reflecting regions present, the heights and sharpness of their lower boundaries, the vertical extent of each region, and the respective fading rates. Records for successive days often differ markedly and do not show a smooth transition from one form to another. Despite this, it is clear that some, if not all, of the factors mentioned may remain constant for as many as four successive days. This grouping of days is quite independent of the similar well-known grouping of days of high absorption. A typical example is that of the daylight hours of July 10–13, during which it was scarcely possible to distinguish a sharp boundary among continuous reflections from 57 km up to the *E* region.

An inspection of solar and magnetic data has revealed no obvious connexions with any group of days, including those of high absorption. Before a more detailed analysis of the separate factors is attempted, it appears desirable to simplify records by eliminating the fading due to interaction between the two magneto-ionic components. Both the variations in each factor and the grouping of days are sufficiently marked to warrant further study.

IV. COMPARISON WITH OTHER RESULTS

(a) *Summary of Other Investigations*

Studies of the lower ionosphere using the pulse method at vertical incidence have been made by Dieminger and Hoffman-Heyden (1952), by Gardner and Pawsey (1953), by Watts and Brown (1954), and by Dieminger (1955). The same method at oblique incidence has been used by Naismith and Bramley (1951), while Gnanalingam and Weekes (1952, 1955) have used a frequency modulation method. In most of these investigations, the frequencies used have been in the range 1–5 Mc/s, and the reflections from below the E region have nearly always been partial. There are also available results obtained with frequencies in the range 16–325 kc/s, for which the majority of reflections are total. A survey of these lower frequency results has been made by Bracewell *et al.* (1951); and later results have been reported by Watts (1952) and by Hopkins and Reynolds (1954).

TABLE 5

SUMMARY OF RECENT RESULTS

Worker	Method	Frequency	Significant Heights (km)		
Dieminger ..	Pulse, vertical	1–5 Mc/s		95, night	75 day (min.)
Gnanalingam & Weekes	Frequency-mod., vertical	1·42 Mc/s	96, day	90, day	74 (min.)
Gardner & Pawsey	Pulse, vertical	2·28 Mc/s		90, day & night	70, day (55 min.)
Gregory ..	Pulse, vertical	1·75 Mc/s	95, day	85, day & night	75, 67, 53 (min.)
Naismith & Bramley	Pulse, oblique	2 Mc/s		90–97, day & night	70–76, day 49 (min.)
Watts & Brown	Pulse, vertical	50–1100 kc/s		85–95 day & night	68–80, day
Bailey, Bateman, & Kirby	Pulse, oblique scatter	28, 50, 108 Mc/s		85–90, night	75–80, day
Bracewell *et al.* (survey paper)	Interference etc.	16–200 kc/s		90–96, night 85, day	70–74, day
Watts	Pulse, oblique	300 kc/s		80–90, night	65–80, day

In considering all these results it is convenient to divide the discussion into two parts, dealing with heights above and below 85 km respectively. Table 5 gives a summary of the main findings of workers in recent years.

A survey of the data summarized in this table shows that agreement between the various results is likely to be affected by the sensitivity and, to some extent, the resolving power of the different methods. It is known that the ionization density in this portion of the ionosphere is basically increasing with height, and therefore an increase in sensitivity is likely to decrease the height at which a given type of irregularity may be detected. This is borne out by the recording of lowest heights with methods of highest sensitivity. Again, many of the regions concerned show negligible group retardation, possibly due to their thinness, and attempts to find critical frequencies on the basis of disappearance of reflection at constant height with varying frequency may be dependent on sensitivity. Times and durations of occurrence are similarly affected. Finally, the ability to differentiate between regions of differing reflection strength and to detect weaker reflections in the presence of stronger has possibly affected a judgment as to whether reflections are total or partial.

(*b*) *Regions above* 85 *km*

At least two regions below the usual *E* region maximum reflection are recognizable. The first, at 95 km, has been described in Section III (*d*) (i) and is also among the reflections noted by Gnanalingam and Weekes. As found in the present work, it is essentially part of the *E* region. The second, at about 90 km during the night, is confirmed by most observers and may be identified with the 85 km region of this paper. Its relation to the phenomenon of very-high-frequency forward scatter has been discussed by Bailey, Bateman, and Kirby (1955), whose results are included in Table 5. The continued existence of this region during the day is also supported ; but its behaviour during certain winter days is not yet clear. A mass plot for 19 days, obtained by Dieminger (1955), shows the same trend of height decrease as Figure 5, which is for one day. In Section III (*f*) (iii) reasons are given for supposing that, during high absorption days, the region retains its identity at 85 km and above, and that the decrease in apparent boundary height is due to the presence of additional ionization below it. Its condition on days of moderate absorption requires further study, since it appears to be involved in the general condition of regions below it.

The origin of this 90 km region appears to be still in doubt. Naismith and Bramley (1951) have suggested that it may be identified with the meteoric region observed by Hey and Stewart (1947). There is certainly a close correspondence between the lowest height at which Hey and Stewart recorded meteor ionization and the 85 km value noted in the present investigation ; but it is difficult to see how the transient meteor occurrences could provide the constant height boundary which is often the chief characteristic of the night 85 km region. Kaiser (1953) has concluded that a continuous band of ionization due to sporadic meteors would form between 130 and 115 km, and that below these heights individual meteors will give only discrete patches of ionization. He also concludes that meteor showers will make a smaller contribution to the ionization continuum than sporadic meteors. During the present series of observations, the night condition of this region has been observed at times of known meteor showers, but appears to be unaffected by them.

Naismith (1954) has drawn attention to the existence of the region and, while maintaining the earlier suggestion of its meteor origin, has considered it as part of the " sporadic E region ". Some of his evidence for this conclusion is based on the " apparently intermittent character " of the region. It seems likely that this intermittency is due to the rapid fading which is often noted at night in present records, and which may have its origin in the existence of separate reflecting or scattering centres. However, the essential feature of present records is the continuity of the boundary at about 85 km rather than an irregularity in time. Further, many instances of sporadic E have been noted, at heights up to 140 km, during the present work, but it is difficult to escape the impression that they have little or no connexion with the 85 km region. An attempt to discover the properties of the region by studying those of sporadic E reflections may thus be misleading.

The model of the night-time D region proposed by Parkinson (1955) on the basis of 150 kc/s data provides for an ionization density which increases from a zero close to 85 km. It is thus attractive as a possible explanation of the night behaviour of the 85 km region, but a closer analysis of present data will be required before the extent of the agreement with this model can be deduced. Finally, it is to be noted that the results of the V.H.F. forward scatter programme in the United States of America indicate that both meteors and continuous ionization are responsible for night-time scattering at heights of 85–90 km. Although meteors have been detected with the present apparatus, they have not been recorded below 100 km, and records such as shown in Plate 2 (C) suggest that essentially continuous ionization is present at about 85 km.

(c) *Reflections below* 85 *km*

Information on reflections below 85 km is less coherent than that for greater heights. Although this situation may be due in part to the weakness of the lower reflections, so that the instrumental differences mentioned in the previous section are aggravated, it seems that the phenomena revealed by the 1–5 Mc/s observations are in fact very complex. This contrasts with the very-low-frequency observations, which are satisfactorily explained by existing theories of single-layer formation.

It is clear that below 80 km there are no regions which compare in permanence and stability with the region lying between 85 and 95 km. There are, however, certain heights, which, on the bases of frequency of occurrence and of strength of reflections, have more significance than others. The first of these is near 75 km, below which value there is, particularly in summer months, a pronounced weakening of reflections. In the present investigation it has been established that the continued recurrence of reflections near 75 km, as shown in Figure 4, is real and not greatly affected by bias in interpretation of records. The second height is near 67 km, from which a winter strengthening of reflections is noticeable. Below this height, the scanty data indicate that values just above 55 km may prove significant. Finally, there is a complete failure to detect reflections below 50 km with the most sensitive apparatus so far employed.

The agreement among all workers as to the appearance during day-time only of reflections below 85 km establishes the solar influence on the lower heights. This serves to emphasize the present lack of lnowledge of the factors responsible for the pronounced increase, as shown in Section III (*e*) (iii), of lowest height reflections during winter time. For this reason, it is not proposed to discuss the relevance of current theories to the results so far obtained. Attractive suggestions have been made by Nicolet (1945) regarding the formation of nitric oxide near 75 km, by Friedman and Chubb (1955) in connexion with possible X-ray sources of the variable ionization at low heights, and by Dieminger (1955) as to the existence of a lower ionosphere meteorology. However, until the information gained from the present investigation can be extended to include quantitative surveys, its bearing on these suggestions is rather limited.

V. Conclusion

This paper presents the interim results of 6 months' operation of a fixed frequency ionosphere recorder of high sensitivity. The work to date has been confined to recording of heights, times, and strengths of occurrence of reflections below the *E* region. From this, it has been possible to deduce the following:

(*a*) The existence of a modification of the *E* region at 95 km, capable at times of total reflection of 1·75 Mc/s pulse transmissions, and showing some measure of solar control.

(*b*) The continuous existence of a region, possibly centred about 90 km and with a lower boundary near 85 km, whose origin is yet uncertain.

(*c*) A direct connexion between the decrease of *E* region reflection strength and the increase in partial reflection strength from a region between 66 and 85 km approximately, on winter days of high absorption.

(*d*) The existence of a complex series of partial reflections, during day-time only, extending down to 53 km, with a marked increase in strength and rate of occurrence at the lowest heights during winter months.

(*e*) The entire absence of detectable reflections below 50 km approximately.

(*f*) Evidence for the existence of a solar ionizing agency which, though variable from day to day in its intensity and penetrating ability, may retain its identity for several successive days.

These results have been compared with those of other workers, with satisfactory agreement in the matter of the 85–95 km region. For lower heights, there do not appear to be sufficient data obtained either at comparable sensitivity or over a sustained period to allow a comparison. Likewise, the present series of results, though giving support to certain current suggestions as to the state of the lowest ionosphere, is not sufficiently quantitative to permit a critical examination of these hypotheses.

VI. Acknowledgments

The work described in this paper has been made possible by the Research Grants Committee of the University of New Zealand, and by the Department of Scientific and Industrial Research of the New Zealand Government, both of which have provided funds for various portions of the project. The author wishes to express his indebtedness to Dr. C. D. Ellyett for helpful criticism and advice during the course of the project.

VII. References

Appleton, E. V., and Naismith, R. (1947).—*Proc. Phys. Soc. Lond.* **59** : 461.

Bailey, D. K., Bateman, R., and Kirby, R. C. (1955).—*Proc. Inst. Radio Engrs., N.Y.* **43** : 1181.

Bracewell, R. N., Budden, K. G., Ratcliffe, J. A., Straker, T. W., and Weekes, K. (1951).—*Proc. Instn. Elect. Engrs.* III **98** : 221.

Dieminger, W. (1955).—" The Physics of the Ionosphere." p. 53. (Physical Society : London.)

Dieminger, W., and Hoffman-Heyden, A. F. (1952).—*Naturwissenschaften* **39** : 84.

Friedman, H., and Chubb, T. A. (1955).—" The Physics of the Ionosphere." p. 58. (Physical Society : London.)

Gardner, F. F., and Pawsey, J. L. (1953).—*J. Atmos. Terr. Phys.* **3** : 321.

Gnanalingam, S., and Weekes, K. (1952).—*Nature* **170** : 113.

Gnanalingam, S., and Weekes, K. (1955).—" The Physics of the Ionosphere." p. 63. (Physical Society : London.)

Hey, J. S., and Stewart, G. S. (1947).—*Proc. Phys. Soc. Lond.* **59** : 859.

Hopkins, H. G., and Reynolds, L. G. (1954).—*Proc. Instn. Elect. Engrs.* III **101** : 21.

Kaiser, T. R. (1953).—*Phil. Mag. Suppl.* **2** : 495.

Naismith, R. (1954).—*J. Atmos. Terr. Phys.* **5** : 73.

Naismith, R., and Bramley, E. N. (1951).—*Wireless Engr.* **28** : 271.

Nicolet, M. (1945).—*Mém. Inst. Mét. Belg.* **19** : 1.

Parkinson, R. W. (1955).—*J. Atmos. Terr. Phys.* **7** : 203.

Watts, J. M. (1952).—*J. Geophys. Res.* **57** : 487.

Watts, J. M., and Brown, J. N. (1954).—*J. Geophys. Res.* **59** : 71.

CUSP TYPE ANOMALIES IN VARIABLE FREQUENCY IONOSPHERIC RECORDS

By G. H. Munro* and L. H. Heisler*

[*Manuscript received May 21, 1956*]

Summary

Anomalous cusps which frequently appear at the high frequency end of ionosonde records of the F_2 region are explained as the result of modification of the ion distribution during the passage of typical travelling disturbances. They indicate the presence, not of vertical stratification but of horizontal gradients of ionization causing oblique reflection. It is suggested that other anomalous cusps are of similar origin.

Anomalies on records of the F_1 region are also shown to be caused by travelling disturbances.

It is demonstrated that these explanations provide useful guidance in the interpretation of ionospheric records and facilitate the further study of travelling disturbances.

I. Introduction

On ionosonde records cusp-shaped traces are normal features indicating the presence of stratification in ionospheric ionization such as the well-recognized E, F_1, and F_2 layers. Quite frequently, however, additional transient cusps are present suggesting additional stratification. They usually appear at the high frequency end of the record and the apparent heights of reflection are greater than that of the normal F region reflections.

Though they are a very common phenomenon by day in temperate regions and they present difficulties in routine scaling of ionospheric records for prediction computation, not much reference has been made to them in published radio literature presumably because there has been no clear explanation of their origin. They are even more evident on records from equatorial stations; for example, Osborne (1951) describes " ledges " with virtual heights of the order of 800 km at Singapore; and Skinner, Brown, and Wright (1954) discuss " ridges " of ionization observed in the F_2 layer at Ibadan.

In Sydney, we have observed large numbers of these cusps on $h'f$ records, and from simultaneous observations on a three-station system for detecting ionosphere movements (Munro 1950) we have established that these transient cusps are always manifestations of ionospheric disturbances travelling horizontally. In the present paper, it is shown that such cusps can be explained as a logical result of a general type of travelling disturbance described in a previous paper (Munro 1950) and due to oblique reflection at the normal height of the maximum ion density. The method of analysis used for this purpose is

---

* Radio Research Board, C.S.I.R.O., Electrical Engineering Department, University of Sydney.

also applied to explain anomalies frequently observed on records of F_1 region reflections. It is suggested that other anomalous cusps may be explained along similar lines, as the result of horizontal ionosphere gradients.

II. F_2 REGION ANOMALIES

Since 1951, a fast-sweeping recorder (Heisler 1955) has been in daily operation at Camden near Sydney, from 0900 to 1400 hr local time. The time of sweep is 15 sec over the range 1–15 Mc/s and records are taken at 1 min intervals. As cusp reflection conditions persist for several minutes this programming ensures that all cusps present are detected and their changes with time can be examined.

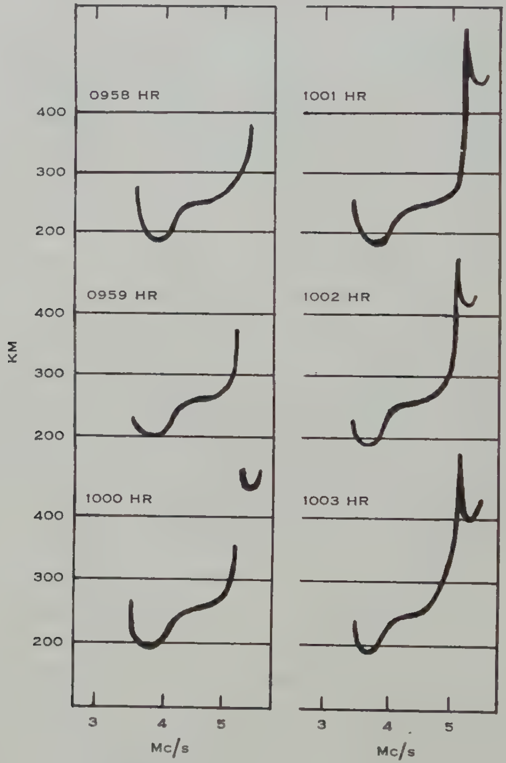


Fig. 1.—Development of a cusp on $h'f$ records during the passage of a travelling disturbance (ordinary ray). Camden, July 31, 1952.

An example of a clear type of simple cusp is shown in Figure 1 which displays a series of tracings of the ordinary ray records. It will be seen that the penetration frequency, which had been constant for several minutes previously, fell appreciably between 0958 and 0959 hr. At 1000 hr an isolated cusp appeared at a greater virtual height and with a penetration frequency greater than that on the first record. In subsequent records, the low frequency end of the cusp merged with the main trace and the virtual height of the bottom of the cusp decreased steadily.

Except for the discontinuity when the cusp appears, the subsequent history of this disturbance is similar to that of disturbances examined in the earlier paper. In that paper true height curves were derived from a set of virtual height curves and the distribution of ionization deduced. The method used there is not fully applicable in the present case because of the discontinuity in the curve. It is possible, however, from optical considerations as applied in a subsequent paper (Munro 1953) to show that, with a general redistribution of ionization such as was found to occur during a typical travelling disturbance, cusp type records would be expected under certain conditions.

The distribution of ionization density in the *F* region during the passage of an actual disturbance as deduced from *h'f* records is shown in Figure 2, where the lines represent surfaces of constant ionization density in a vertical cross section in the direction of travel of the disturbance. It will be noticed that at any point, the disturbance causes a decrease in ionization density followed by an increase, resulting in a curvature in the contours which has a relative displacement horizontally (or change of phase) with height. It should be noted also that while the height of maximum ion density remains constant, the value changes, being a maximum in the compression phase towards the left and a minimum in the expansion phase to the right, as indicated by arrows on the base-line.

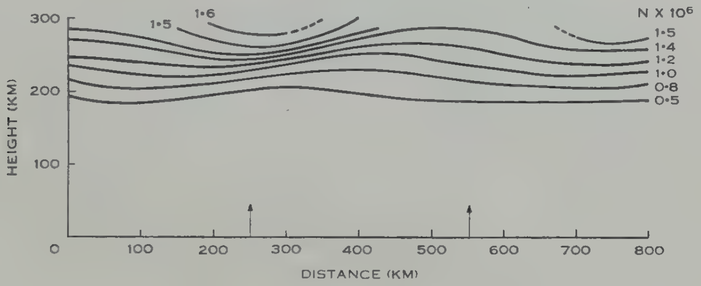


Fig. 2.—Cross section of a travelling disturbance in the *F* region of the ionosphere. (Lines represent isoionic surfaces or surfaces of equal ionization.)

In earlier papers on this subject, the discussion was mainly in terms of reflection patterns as observed on records, but it is clear that variations in the reflected ray paths and movement in the points of reflection must occur. Such variations have been confirmed by direction-finding observations particularly by British Radio Research Board investigations (e.g. Bramley and Ross 1951).

The concave reflecting surfaces in the reduced ion-density phase are mainly responsible for complexities in the reflected rays, so in the earlier paper (Munro 1953), in considering complexities, a synthetic diagram was constructed with curves of similar form to those actually observed (Fig. 2) but of known curvature which permitted easy geometrical treatment. On such a model, it is possible to locate readily the reflection points for all rays which retrace their paths. A synthetic diagram of this type corresponding approximately with the observed conditions in Figure 2 is used in Figure 3 to indicate the changes in returned ray paths as a disturbance passes over an observing point on the ground. The series of lettered points along the base-line will, therefore, represent positions of the observing point relative to the disturbance at successive equal intervals of time. The straight lines indicate the paths of rays which are reflected back to the sending point from the lowest surface shown (surface 1). It will be seen that the path deviates from the vertical first to the right, the deviation increasing steadily to a maximum, then sweeps back rapidly to an equal maximum deviation to the left, and, finally, slowly reverts to vertical. The same cycle is followed for the other surfaces but later in time. The points connected by dotted lines represent the reflection points at the four surfaces for a particular observing site at the ground so that the dotted lines would approximate to the loci of the reflection points as a recorder at the corresponding site swept through a range of frequencies.

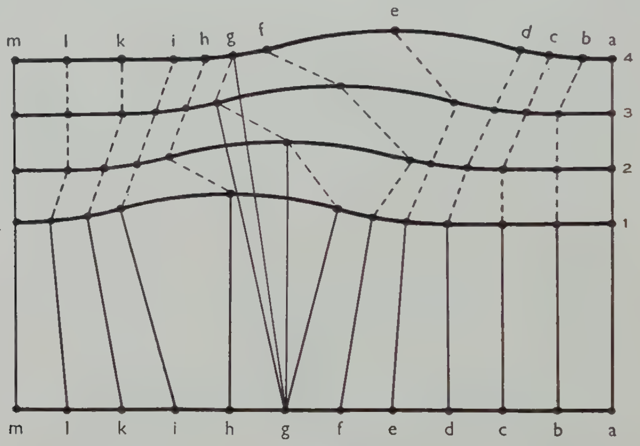


Fig. 3.—Displacement of reflection points by a travelling disturbance under simple reflection conditions.

In this case, there are no complex reflections but the ray paths change as the frequency is varied and these changes would vary with time as the disturbance passed overhead. This change in ray path with frequency would, therefore, introduce errors into the estimation of true heights from *h′f* curves taken under corresponding conditions. Some error of this nature would be present in Figure 2, but the only effect there would be to distort the shape of the curves slightly which would not in any way affect the deductions made therefrom.

In Figure 2 the ionization density was seen to be greatest at the left. It may be seen also in Figure 3 that the reflection from surface 4 observed at point *f* is in the corresponding part of the diagram. This displacement of reflection points becomes more pronounced as the curvature is increased. Figure 4 is similar to Figure 3 except that the curvature is now greater and will give complex reflection from all surfaces above the lowest shown. Again the dotted lines show the loci of reflection points as the frequency is changed.

In this case, for reflections from surface 1, marked focusing would be observed at point *h*. The higher surfaces will give complex reflections with rays returned from two or three points. For example, it will be noticed that for the frequency corresponding with surface 4 there will be, at point *d*, three returned rays reflected from widely spaced points, d^1, d^2, and d^3, with noticeably different ion distribution in the paths. Moreover, if the maximum density of ionization were lower at the right in Figure 4, say, that of surface 3, the minimum frequency returned from the region d^2, d^3 could be greater than the maximum reflected in the region near d^4, thus giving rise to an isolated cusp on a record as in Figure 1.

At observing points further to the left, the cusp and the main trace would tend to merge or overlap until at *g*, where a complex reflection of the loop or distorted loop type would appear, as shown in a previous paper (Munro 1953).

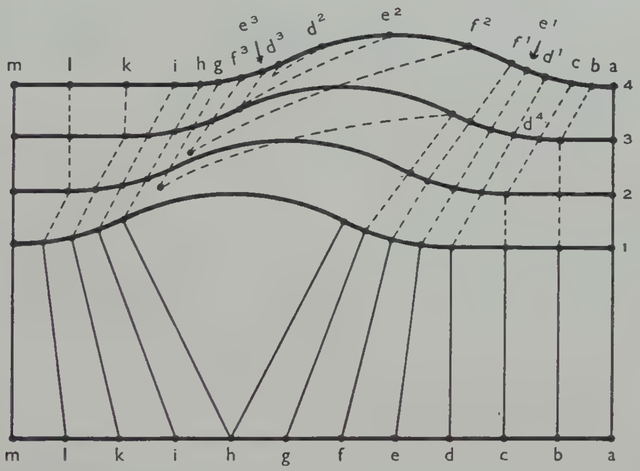


Fig. 4.—Displacement of reflection points by a travelling disturbance under complex reflection conditions.

In Figure 1 there is no clear complexity after the two traces join. This implies a rapid transition from the reflection conditions of surface 4 in Figure 4 to those of surface 1. At point *h*, strong echoes would be returned from this lowest surface from a wide range of angles and with a range of heights and retardation. These features are apparent at 1001 hr in Figure 1.

It was pointed out that, for the conditions in Figure 4 where the curvature is constant, complex reflections would be returned from all heights above surface 1 but not below that surface, where reflection paths would be similar to those of Figure 3. However, if the curvature were to decrease below surface 4 the change to simple reflection would occur sooner. There is some indication in Figure 2 that the curvature does decrease from the top downwards. It seems likely that this is a result of the changing ionization gradient. This possibility is examined by an indirect approach in Section III.

III. CHANGES IN THE DISTRIBUTION OF IONIZATION DUE TO TRAVELLING DISTURBANCES

From the discussion in Section II, it can be seen that there are limitations to a direct method of analysis of these disturbances. It is not possible to determine vertical ionization distributions by an integration process (e.g. Manning 1947) because of the discontinuity in the curves and the probability that the path of the returned ray will vary with the frequency, as indicated in the previous section. In addition it is not possible by reflection methods to determine the distribution of ionization above maximum ion density.

It is possible, however, to obtain a more complete picture of the probable changes by an indirect approach. Where the changes can be observed and interpreted clearly, the redistribution of ionization is such as would result from a travelling atmospheric disturbance whose characteristics have been defined previously. It is, therefore, reasonable to assume that the characteristics of this disturbance are the same throughout the region. Its effect on the ionization throughout the whole region may then be deduced by starting with a normal undisturbed ion distribution and studying the changes in distribution as the disturbance progresses.

For this purpose it has been assumed that:

(i) the disturbance consists of a single oscillation of sine waveform of such a nature as to change the concentration and therefore the distribution of ionization;

(ii) the disturbance travels horizontally with its front tilted forward at an angle of 45°; and

(iii) the disturbance travels so that the trough of the assumed compression wave precedes the crest. This is always observed in practice as a reduction in critical frequency followed by an increase.

This concept is considered to be a reasonable simplification based on observed characteristics of the disturbances and on the type of ion redistribution they produce without any assumptions as to the details of the propagation mechanisms involved.

The effect of such a disturbance on the vertical gradient of ionization has then been calculated for a series of epochs by multiplying the undisturbed values by the appropriate disturbance factors, and deriving contours of constant ionization density from these modified values. We thus get a representation of a vertical cross section of the disturbance in the direction of travel. As the ionization gradient above the maximum is not known, it is assumed to be symmetrical about the maximum. An example derived in this way is shown in Figure 5. The assumed form of the disturbance and the undisturbed ionization gradient are shown at the left of the diagram.

The modified distribution at the height of maximum ionization density is clearly shown, the main features being a " trough " of reduced ionization followed by a " ridge " of increased ionization having a localized maximum peak or " island " at its centre. Both trough and ridge are inclined to the vertical. Since a travelling disturbance is a frontal phenomenon, this island really represents a long horizontal pencil of high ionization.

It will be seen that the shapes of the contours below the maximum are very similar to those in Figure 4. It will be noticed also that the curvature does decrease at lower heights, and it is asymmetric. These features are conducive to the production of a cusp type of record as indicated in Section II, the cusp being due to reflections from the tilted " island " of high ionization density. From examination of ray paths it may be deduced that retardation will be initially high for these reflections. That the great apparent height of the cusp in Figure 1 is mainly due to retardation has been confirmed by phase-path records

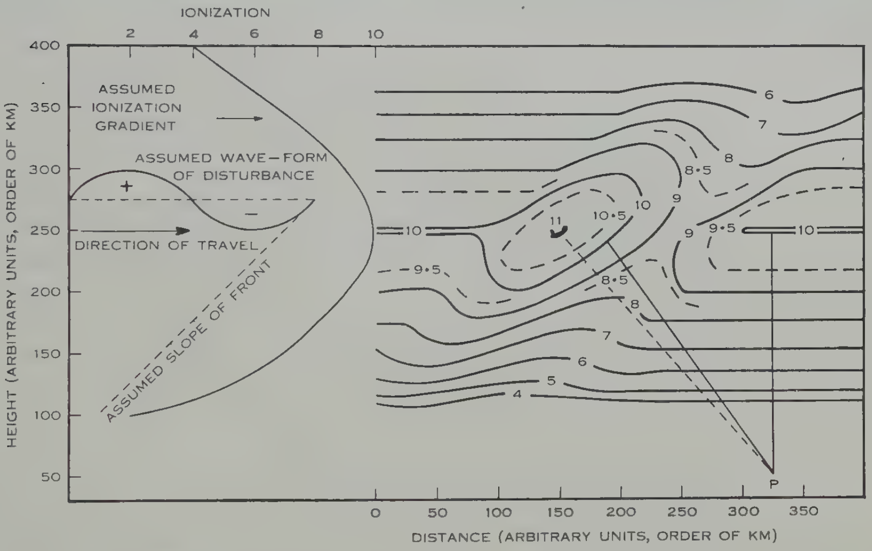


Fig. 5.—Distortion of ionization contours by a travelling disturbance. F_2 region. (Ionization units chosen to correspond approximately with normal frequencies of reflection.)

taken by a method previously described (Munro 1953). In a typical case, where the virtual height of a descending cusp changed by 150 km, the phase-path change was of the order of only 15 km.

In scaling ionospheric records, it has been the custom to regard such cusps as abnormal stratification at a high level (occasionally called a G layer or F_3 layer), and to neglect them in scaling F_2 critical frequencies. It would now appear more correct to scale the maximum frequency of the cusp, as it is a true maximum of F_2 occurring at the normal height.

In winter, when these cusps are most frequently observed, the average rate of occurrence at Sydney is at least one per hour between 1000 and 1400 hr local time. The percentage difference between maximum ion density as measured from the cusp and the maximum ion density of the normal F_2 region is on the average 20 per cent. and may be as high as 40 per cent. The ambiguity is present for about 5 min for each disturbance. Reading the cusp penetration frequency would therefore give at this time of day a mean increase of the order of 2 per cent. in maximum ion density over that deduced using present methods of interpretation.

IV. F_1 Region Anomalies

The distribution of ionization in the F_1 region also is frequently modified by travelling disturbances. The large winter disturbances, which first become evident on $h'f$ records at the top of the F_2 region as in Figure 1, can generally be observed to progress downward through both regions; but many disturbances, particularly at the equinoxes and in summer, cause noticeable distortion of the curves only in the F_1 region. This takes the form of an increase in virtual height causing an accentuation of the normal F_1 peak sometimes accompanied by complexities near the peak.

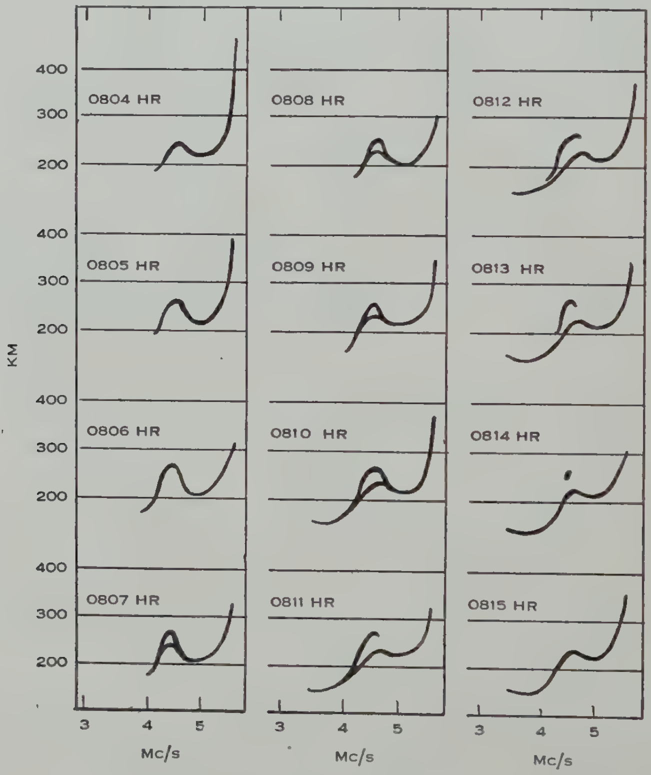


Fig. 6.—Effect of a travelling disturbance in the F_1 region on $h'f$ records (ordinary ray). Camden, October 10, 1952.

An example of an F_1 complexity is shown in Figure 6, which again consists of tracings of the ordinary ray from a series of records. A complete F_1 penetration does not occur under normal conditions in Sydney, but there is frequently an F_1 peak in the $h'f$ curve as shown in the first curve in Figure 6. It will be seen that the height of the F_1 peak increases after 0804 hr and then, in the tracing for 0808 hr, a second smaller peak appears below the original one. This new peak then becomes the continuing one and the older peak becomes discontinuous and finally disappears. It will be seen also that the two peaks occur at slightly different frequencies.

Although the distortion in the curve here is very different in appearance from that in Figure 1, it may still be the result of a similar disturbance, the difference in the distortion of the curve arising from the different initial ion distribution in the two regions. This may be demonstrated using the method of Section III. Since there is not a complete penetration in this case the ionization gradient will not actually reverse, but will decrease in the F_1 region. In Figure 7, we assume an undisturbed ionization gradient of this type and examine the changes in the distribution of ionization which would be caused by a travelling disturbance of the same characteristics as in Figure 5. The resulting contours of constant ionization are shown at the right.

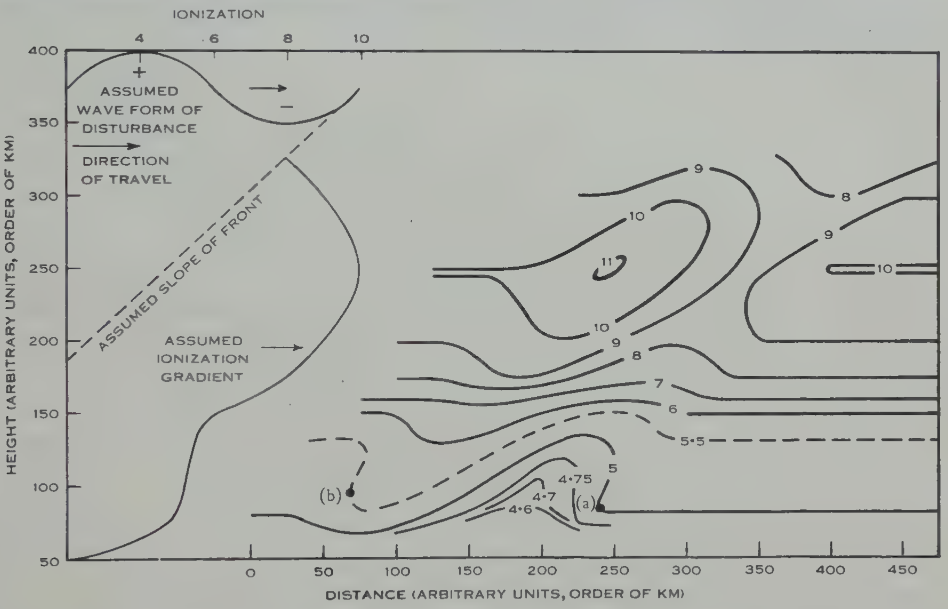


Fig. 7.—Distortion of ionization contours by a travelling disturbance in the F_1 region.

It will be noticed that the curvature decreases from the F_2 maximum downwards but increases again in the F_1 region, thus providing conditions favourable to complex reflection. It will be observed also that the curvature is markedly asymmetrical, which will tend to give a complexity with two ray paths in stead of three as was shown to be the case with a symmetrical " kink ". A study of ray paths shows that a distribution such as this can explain the type of complexity shown in Figure 6.

It is not proposed to give here a more detailed study of the possible variations in patterns, but it will be seen that, while the ion contours will vary with changes in the ionization gradient and disturbance characteristics, the main features will be as shown in Figure 7.

It should be noted that the sharpness of the bends in the contours at points (*a*) and (*b*) in Figure 7 is due in part, but not entirely, to the assumption that the form of the disturbance is one cycle of a sine wave oscillation. It is unlikely that a disturbance would actually start and end so abruptly.

Here again, a knowledge of the nature of travelling disturbances and their effects assists in the interpretation of records. As the result of a disturbance the F_1 peak on an *h′f* record will generally be accentuated and a peak, a complexity, or even an apparent penetration may appear temporarily, which would normally be read as an indication of an F_1 critical frequency. Examples of such anomalies were shown in previous papers (Munro 1949, 1950). It is possible that when disturbances appear only in the F_1 region, it is because the ionization gradient is such that a given disturbance produces the maximum distortion of contours in that region.

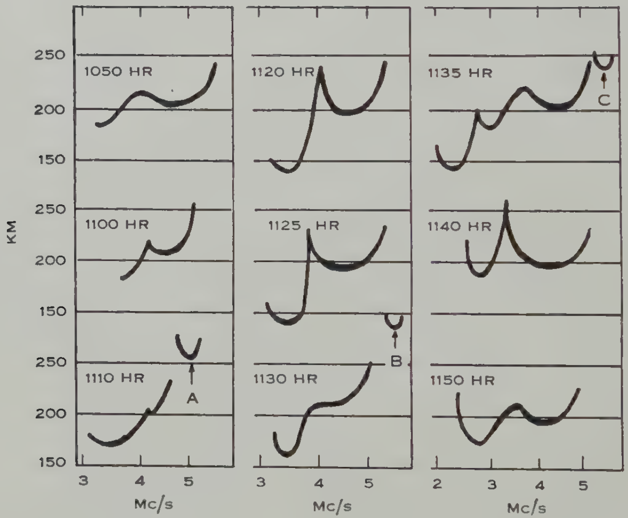


Fig. 8.—Travelling disturbances affecting both F_2 and F_1 traces. Camden, July 31, 1953.

When disturbances are frequent the variation in *h′f* records may become very complex. A sequence of records of the ordinary ray taken under such conditions over a period of 1 hr is shown in Figure 8.

The conventional interpretation would probably be as follows : At 1050 hr the record is normal, the F_1 region appearing as a ledge of the F_2 and not a complete layer. At 1100 hr the F_2 critical frequency has fallen and the F_1 region is almost a separate layer. At 1110 hr the F_2 region is bifurcated, giving an extra layer *A* at a greater height. At 1120 and 1125 hr, the F_1 has a definite critical frequency which is different in the two cases. At 1130 hr the F_2 is again bifurcated while at 1135 hr it is triply stratified. At 1140 hr the F_1 has a definite critical frequency.

The explanation in terms of disturbances is quite straightforward : At 1100 hr the F_2 critical frequency has fallen due to the advent of a disturbance.

At 1110 hr the disturbance has caused a cusp to appear at A and the main curve is distorted. At 1120 and 1125 hr the cusp has merged with the main curve and the disturbance is accentuating the F_1 peak at progressively lower frequencies. At 1130 hr disturbance A has passed on and F_1 is almost normal; but a second disturbance causes a new cusp B. At 1135 hr the cusp of disturbance B has joined the main F_2 curve and a third disturbance C appears as a new cusp. At 1140 hr disturbances B and C are both affecting the F_1 region.*

V. Applications

It is considered that Figures 5 and 7 show the changes in distribution of ionization during the passage of a travelling disturbance sufficiently well to enable them to be applied in the interpretation of records for a number of purposes, for example:

(i) We may with advantage use such a model as a guide in the interpretation of effects on records to improve the accuracy of extraction of parameters for statistical purposes; e.g. in reading $f^\circ F_1$ and $f^\circ F_2$ as indicated.

(ii) This knowledge of the nature of disturbances on $h'f$ records may be applied to the study of the divergence of the ordinary and extraordinary rays. This application is described in a separate paper (Munro and Heisler 1956).

(iii) Such a model is useful in investigating the incidence of a disturbance over a range of heights as it enables each phase of a disturbance to be recognized in its varying forms on an $h'f$ record. This is of considerable importance in tracing apparent vertical progression as some disturbances appear to progress right down to the E region. This also is being studied.

(iv) The anomalies discussed here occur particularly in the neighbourhood of Sydney (lat. 33° 52′S., long. 151° 11′E.), and the understanding of their nature makes possible the study of similar effects on $h'f$ record from other localities. Records from Hobart (42° 53′S., 147° 51′E.), Canberra (35° 18′S., 149°E.), Brisbane (27° 30′S., 153°E.), Watheroo (30° 18′S., 115° 56·6′E.), and Townsville (19° 10′S., 146° 58′E.) have been examined in this way and found to show disturbances similar to those seen in Sydney; records from other countries have been found to show similar phenomena. In addition, evidence has been found which suggests that these disturbances travel unchanged with very wide fronts over considerable distances. This study is being continued.

(v) It was mentioned earlier that radio direction-finding errors have been observed during the passage of travelling disturbances. The reason for such variations can be understood readily from examination of

---

* *Note added in Proof.*—Actual confirmation of the correctness of this interpretation is provided in a recent paper (Jackson 1956) on rocket observations where the rocket apparently passed through a typical disturbance. The $h'f$ there shows cusps whereas the rocket recorded gradual changes but no stratification.

Figures 5 and 7 in conjunction with Figures 3 and 4. Direction-finding observations would, in turn, be of use in the more detailed study of disturbances. It should be noted that the effect of refraction has been neglected in the present analysis but could be appreciable in directional observations.

(vi) Since the distribution of ionization during a disturbance seems satisfactorily established, a basis is provided for considering its effects on long distance propagation, and propagation through the ionosphere as in radio astronomy.

(vii) These models may be used to test theories of the formation and mode of travel of disturbances. For a theory to be satisfactory, it must be able to explain the production of ion distributions similar to those considered above. There are theoretical objections to the simple mode of propagation of a disturbance assumed here. These were discussed by Martyn (1950), who put forward a cellular wave theory to avoid them. This theory, however, does not appear to satisfy all the facts. For example, the slope of the front of a disturbance is explained as an effect of the Earth's magnetic field. If this were so, we would not expect any slope if the disturbance were travelling across the field, but our indications are that there is still a slope under these conditions. Also, in the cellular wave theory, one would expect the maximum of ionization density to change in height, but not in intensity, whereas our indications are that it changes in intensity but not in height.

VI. Discussion of Other Reported Anomalies

We have considered various abnormal patterns in ionospheric records and have shown how they may be interpreted in terms of changes of distribution in ionization due to travelling disturbances.

It seems necessary to emphasize that, in our extensive studies of ionospheric disturbances, we have always found evidence that these disturbances move horizontally. There is also present, in all cases we have checked, an apparent vertical component of movement which we attribute to a tilt in the front of the disturbance. The most obvious result is a varying tilt of the effective reflecting surface for a radio wave and a consequent displacement of the reflection point from vertically overhead. These effects have been confirmed by direction-finding observations (Bramley and Ross 1951).

In the above discussion on travelling disturbances we have been considering a particular case of a general principle which may be stated as follows: If a horizontal gradient of ionization density exists at any point in the ionosphere then the surfaces of equal ionization (or isoionic surfaces) will no longer be horizontal and the corresponding reflection points for so-called " vertical incidence soundings " will no longer be vertically overhead.

Horizontal gradients may arise from redistribution of ionization as in the case of a travelling disturbance or from horizontal progression of an increase or decrease in ionization density such as the normal diurnal variation.

If either the initial vertical gradient or the rate of change is not linear the surfaces will be curved and rays will be focused or dispersed causing intensity changes to be recorded, and, if the curvature exceeds a critical value, reflections will be complex.

In addition, changes in gradient along the ray paths will cause variations in observed retardation and absorption. These possibilities must all be taken into account when considering anomalies on records. The following examples of anomalies reported elsewhere are of particular interest.

(*a*) *Sunrise Cusps*

Cusps similar to those we have considered have been noticed near sunrise (Bhargava 1952), but they change more slowly and tend to recur at the same time on successive days. It seems probable that they are the result of non-vertical reflections rather than separate layer formation, as the progression of the Sun's ionizing radiation must result in tilted reflection surfaces. As the Sun rises the ionization at all levels increases. Thus just after sunrise at any point in the ionosphere the isoionic surfaces will dip towards the east. We then have a distribution of ionization similar to the centre portion of Figure 2 between the arrowheads. There will therefore be displacement of reflection points ; and, if the change of ionization is rapid, as would occur at low latitudes, a cusp or complex reflection could result. This displacement could be examined by direction-finding observations.

The presence of such tilts has in fact been demonstrated by direction-finding techniques by Bramley (1953) and Whale (1955).

The change of height of isoionic surfaces and, therefore, the curvature will be greatest where the vertical gradient of ionization is least ; that is, near the maximum of ionization.

(*b*) *Solar Eclipse Anomalies*

In discussion of radio observations during solar eclipses little consideration seems to have been given to the possibility of displacement of reflection points from overhead.

The eclipse shadow may be regarded as approximating to a series of concentric cylinders, graduated from maximum to zero sunlight, generally inclined to the vertical, and progressing mainly from east to west through the ionosphere. At any point in the ionosphere the ionizing radiation will therefore gradually fall to a minimum and increase again to a relatively steady value. The density of ionization will change correspondingly at a rate depending on the ionizing and recombination processes, which will vary with height.

If, therefore, we look at the spatial situation at any instant the resultant redistribution of ionization will be very similar to that caused by a travelling disturbance except that the eclipse disturbance is circular rather than frontal. The resulting contours will appear in cross section very like those in Figures 5 and 6. We would expect, therefore, anomalies in both the F_2 and F_1 regions.

In discussions of eclipse results it has been tacitly assumed that ionosondes record vertical soundings. While this will be sensibly true for the E region it will certainly not be so for the F_1 and F_2 regions. Owing to the curvature of the isoionic surfaces it will, in fact, be very difficult to record the penetration frequency at the centre of the shadow, as rays will be returned obliquely from the regions of higher ionization to the sides as in Figures 3, 4, and 5. The loci of reflection points for a circular kink were discussed in an earlier paper (Munro 1953), and it was there shown that there is no reflection from the centre unless the observing point is directly below it. In an eclipse this will only apply for one height for a very short time if the shadow is not vertical. Moreover, the reflection point moves rapidly at the peak of the curve in the reflecting surface so that it would require records at very frequent intervals to be sure of recording at the peaks.*

In most observations discussed to date, records have been made at 4–10 min intervals, requiring much interpolation. In a comprehensive paper by Minnis (1955) describing eclipse observations at Khartoum, it is shown that a flattening of the minimum of ionization occurs in the F_1 and F_2 regions—in other words, the observed reduction of ionization is less than expected. In the paper, this is ascribed to active regions on the limbs of the Sun. Optical effects as discussed in the present paper would, however, have similar results to those observed as, owing to the curvature, reflections would be observed from the more highly ionized regions surrounding the central minimum. It should be noted also that the $h'f$ curves in Figure 9 of the Minnis paper show distortions very similar to those in Figure 8 of the present paper, which could arise from changes in gradient due to the eclipse itself possibly augmented by the presence of travelling disturbances.

In the E region the distortion of surfaces will be smaller but may still be sufficient to account for the asymmetry observed.

Detailed examination of the full observational data would be required to check these effects quantitatively.

(c) *Equatorial Cusps*

(i) *Huancayo.*—Anomalies in the F_2 region observed at Huancayo may also have a similar explanation. $h'f$ records from this station during the morning hours at sunspot minimum show a spur (similar to " cusps " mentioned above) at the high frequency end. This " spur " moves to higher frequencies and finally sometime before noon disappears at the top of the trace and is no longer seen. Ratcliffe (1951) analysed these records to obtain n, the total ion content of the layer, and found that it was necessary to include the " spur " in estimations in order to obtain consistent results. Moreover, upon disappearance of the spur results became inconsistent, the values of n obtained being too low. Martyn (1955) has published average world contour maps of $f°F_2$ at the magnetic equator which passes through Huancayo and these suggest that the spurs may be off-angle reflections from long ridges of high ionization density north and south of the trough. Disappearance of the spur could result from geometric changes in the reflecting ion density surfaces such that a true vertical estimate of the ionic content of the region would not be obtained, as for example just before a cusp appears in Figure 1.

---

* *Note added in Proof.*—In a recent paper (Bramley 1956) the presence of varying tilts during an eclipse was actually recorded.

(ii) *Singapore and Ibadan*.—Published reproductions of *h'f* records from equatorial stations such as Singapore and Ibadan (Osborne 1951; Skinner, Brown, and Wright 1954) show multiple layer effects very similar to those of travelling disturbances. The tendency so far seems to have been to attempt to explain them in terms of horizontal stratification or vertical movements. It seems probable, however, that the high vertical heights and rapid vertical drift of ionization reported at these places can also be satisfactorily explained in terms of disturbances travelling horizontally in the ionosphere. Observations at spaced points are necessary to check this.

VII. Conclusions

Changes in apparent height of reflection of radio waves as normally recorded with either change of frequency (*h'f* records) or lapse of time (*h't* records) are mainly the result of variations in the vertical gradient of ionization.

The present analysis shows, however, that similar changes may result from the presence of a varying horizontal gradient of ionization. In this case, there are generally also deviations from a vertical path of rays transmitted and received at adjacent points.

Some typical results are shown of anomalies on records due to horizontal gradients caused by travelling disturbances and it is indicated that certain other anomalies observed elsewhere may have a similar explanation.

It is therefore important that, in the interpretation of ionospheric records, the possibility of distortions due to horizontal gradients be considered carefully.

VIII. Acknowledgments

The detailed reduction of records and preparation of material by the analysis staff of the Radio Research Section, Sydney, are gratefully acknowledged. The authors express their appreciation to the University of Sydney and in particular to Professor D. M. Myers for provision of facilities in the Electrical Engineering Department.

IX. References

BHARGAVA, B. N. (1952).—*Nature* **170**: 983.
BRAMLEY, E. N. (1953).—*Proc. Roy. Soc.* A **220**: 39.
BRAMLEY, E. N. (1956).—*J. Atmos. Terr. Phys.* **8**: 98.
BRAMLEY, E. N., and ROSS, W. (1951).—*Proc. Roy. Soc.* A **207**: 252.
HEISLER, L. H. (1955).—*Aust. J. Appl. Sci.* **6**: 1.
JACKSON, J. E. (1956).—*J. Geophys. Res.* **61**: 107.
MANNING, L. A. (1947).—*Proc. Inst. Radio Engrs. N.Y.* **35**: 1203.
MARTYN, D. F. (1950).—*Proc. Roy. Soc.* A **201**: 216.
MARTYN, D. F. (1955).—"Physics of the Ionosphere." p. 260. (Physical Society: London.)
MINNIS, C. M. (1955).—*J. Atmos. Terr. Phys.* **6**: 91.
MUNRO, G. H. (1949).—*Nature* **163**: 812.

Munro, G. H. (1950).—*Proc. Roy. Soc.* A **202** : 208.
Munro, G. H. (1953).—*Proc. Roy. Soc.* A **219** : 447.
Munro, G. H., and Heisler, L. H. (1956).—*Aust. J. Phys.* **9** : 359.
Osborne, B. W. (1951).—*J. Atmos. Terr. Phys.* **2** : 66.
Ratcliffe, J. A. (1951).—*J. Geophys. Res.* **56** : 487.
Skinner, N. J., Brown, R. A., and Wright, R. W. (1954).—*J. Atmos. Terr. Phys.* **5** : 92.
Whale, H. A. (1955).—*Nature* **175** : 77.

DIVERGENCE OF RADIO RAYS IN THE IONOSPHERE

By G. H. Munro* and L. H. Heisler*

[*Manuscript received May* 21, 1956]

Summary

Travelling disturbance manifestations on ionosonde records usually occur at different times on the " o " and " x " traces. It is shown that this is due to the divergence of the ordinary and the extraordinary rays in the Earth's magnetic meridian plane and that the sense of this time difference, therefore, gives a direct indication of the sense of the north-south component of movement of the disturbance. Furthermore, where the direction and speed of travel of a disturbance can also be determined from spaced station observations, the actual separation of reflection points of the o- and the x-rays can be deduced. At Sydney, N.S.W. (33° 52′ S., 151° 11′ E.) this is of the order of 30 km at the height of maximum ionization of the F_2 region. On $h't$ records the corresponding time difference observed includes a component due to the vertical separation of reflection points if the front of the disturbance is not vertical. From records of this type taken at three spaced stations the horizontal component may be determined directly. Examination of some 430 such observations taken over a period of 4 years confirms the variation of the time difference with direction of travel predicted by theory, and also indicates the consistent presence of a forward tilt in the front of disturbances.

The theoretical relation between the horizontal and vertical separation is then used to determine the slope of front of disturbances. It is found to have a mean value varying from 65° to the horizontal for disturbances travelling northward to a minimum of 51° for those travelling 120° E. of N.

I. Introduction

In previous papers by one of the authors (Munro 1949, 1950, 1953), the main characteristics of travelling disturbances in the F_2 region of the ionosphere were deduced, and resulting effects commonly observed in ionospheric records explained. It has been observed on both $h'f$ and $h't$ records that these disturbances generally affect the ordinary and the extraordinary rays similarly but at different times. This must be due to an actual spatial separation of the reflection points of o- and x-rays for reflections observed at the recording site.

Booker (1938), Millington (1949), and more recently Kirkpatrick (personal communication) have shown that in the presence of the Earth's magnetic field, magneto-ionic splitting of a radio ray occurs and the ordinary and the extraordinary rays diverge from the vertical in opposite directions in the Earth's magnetic meridian plane, so that the reflection points of these rays are spaced both horizontally and vertically in the ionosphere. This separation has been

---

* Radio Research Board, C.S.I.R.O., Electrical Engineering Department, University of Sydney.

calculated theoretically by Kirkpatrick for any magnetic latitude and his results are shown in Figure 10.

In general, a difference between arrival times of a disturbance at the ordinary and the extraordinary reflection points may be due to both horizontal and vertical progression of the disturbance front. This progression is mainly horizontal, since observations using a three-point network (Munro 1950) establish definitely an apparent horizontal component of motion, and, therefore, since the extraordinary ray deviates to the north in the southern hemisphere and to the south in the northern hemisphere, order of arrival of a disturbance at the *o* and the *x* reflection points indicates immediately the general direction of travel of such disturbances.

It has also been observed that disturbances travelling in a direction normal to the magnetic meridian plane arrive at different times at the *o* and the *x* reflection points. This time difference can only be due to an apparent vertical component of motion of the disturbance, and supports an earlier assumption in a previous paper (Munro 1953) that the disturbance has a tilted front.

In the present paper observations of time differences between arrival of disturbances on *o* and *x* traces on *h'f* records are used to determine the horizontal separation between *o* and *x* penetration points in the F_2 region. Similar time differences, obtained from three-point fixed frequency observations, together with speeds and directions, are used to check divergence theory and to measure the tilt of the front of the disturbance. Theoretical values for separation deduced by Kirkpatrick as shown in Figure 10 are also used to study travelling disturbances.

II. Variable Frequency Records

Since 1951, at Sydney (33° 52′ S., 151° 11′ E.) regular *h'f* recordings have been taken at 1 min intervals, with a fast sweeping recorder previously described (Heisler 1955). These records show many examples of cusp type anomalies such as that shown in Figure 1. This is a series of *h'f* records taken at 1 min intervals, showing a disturbance which first becomes evident in the form of an isolated cusp appearing at or close to the normal penetration frequency and at a relatively great height. In our records, this height is usually of the order of 500–600 km. On successive records the cusp moves steadily downwards and eventually joins the main curve. The cusp generally appears first on the ordinary ray trace and later on the extraordinary ray. The origin of these cusps is explained in an accompanying paper (Munro and Heisler 1956).

The ordinary and extraordinary rays at a given frequency f deviate in the F_2 layer as shown in Figure 2 and are reflected at A and B, two points separated both vertically and horizontally in the magnetic meridian plane. There is a frequency f_{ex} greater than f, the extraordinary ray of which deviates as shown by the dotted line in Figure 2, and is reflected at B', which is at the same height as A. This may be defined as the " equivalent frequency ", or the value of frequency at which the extraordinary ray is reflected at the same height as the ordinary ray of a given frequency.

The *o*- and *x*-ray frequencies which are reflected at any particular height will be equivalent frequencies, but during the passage of a disturbance, because of the horizontal separation of the reflection points, they may be returned from this height at different times.

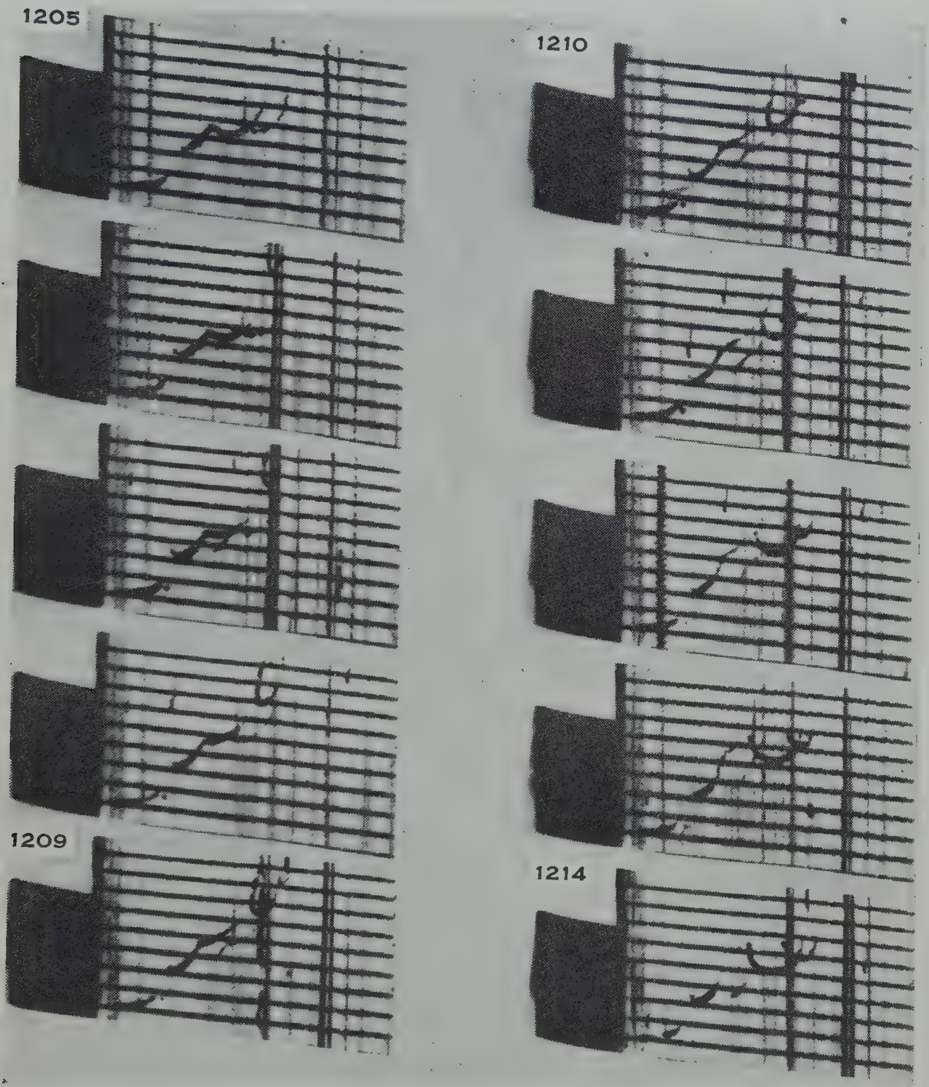


Fig. 1.—A series of $h'f$ records showing a cusp type anomaly.

It follows, therefore, that, if we can detect an indication of a disturbance on the $h'f$ curves at a frequency f on the ordinary ray, and a frequency f_{ex} on the extraordinary ray, then a knowledge of time difference of occurrence at these points, together with the known speed of travel and direction of travel from three-station fixed frequency observations, enables us to estimate the horizontal separation of the two reflection points AB'.

In the absence of disturbances, the o- and x-rays must penetrate the layer at the same height, the height of maximum ionization. For frequencies not far from the penetration frequency, therefore, we may assume that the difference between f and f_{ex} is the difference in penetration frequencies.

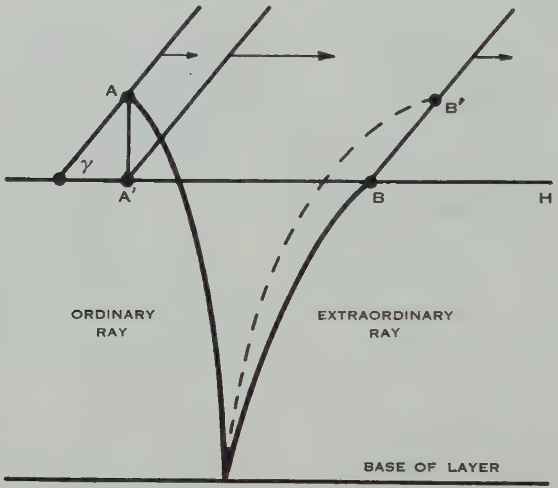


Fig. 2.—Deviation of o- and x-rays in the ionosphere.

Referring to Figure 1, at the bottom of the cusp the rate of change of height with frequency is small so that the equivalent frequencies need not be accurately determined. We may simply assume that on the records they will be at the bottom of the respective cusps at the same height.

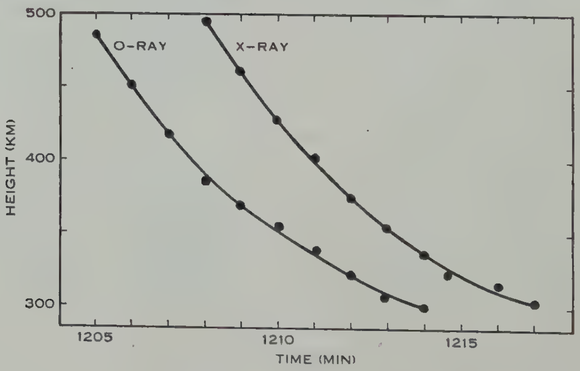


Fig. 3.—Variation with time of height of cusps observed on o- and x-rays.

If, therefore, we plot against time, the heights of the lowest points of the two cusps observed on each record over a period of several minutes, and join the points, we obtain two similar curves with a time displacement. The results obtained in this way for the disturbance of Figure 1 are shown in Figure 3.

It will be seen that the virtual height change is almost linear in the first few minutes and the time displacement is reasonably constant at 3·5 min. It has been shown elsewhere (Munro and Heisler 1956) that, although the virtual height change of a cusp is great (of the order of a hundred kilometres), the actual height change is much less (of the order of tens of kilometres) so that this constancy is not unexpected.

The direction of horizontal movement in this case as obtained from the three-station observations was 8° E. of N., which is almost the declination of the magnetic field, so no correction is necessary. The horizontal speed was 9·8 km/min, the separation of reflection points would, therefore, be $3 \cdot 5 \times 9 \cdot 8 = 35$ km approximately. This would be the horizontal separation near the maximum of ionization. The results for a number of similar cases are given in Table 1.

TABLE 1

HORIZONTAL SEPARATION BETWEEN THE ORDINARY AND THE EXTRAORDINARY RAY REFLECTION POINTS FOR A NUMBER OF CASES OF TRAVELLING DISTURBANCES

Date	Time	Time Difference, t $(o—x)$ (min)	Direction θ	Speed S (km/min)	Separation (km) $(t \times S \cos \theta)$
18. vi.52	1020	3·6	353°	8·8	31
10. vii.52	1142	3·0	26°	9·4	25
8.viii.52	1342	3·6	325°	10·8	31
26.viii.52	0930	4·3	3°	6·9	30
	1205	3·25	8°	7·8	32
	1310	3·5	355°	10·0	35
27.viii.52	1140	3·6	26°	9·4	30
					Mean : 30·5

It should be noted, that observation of cusps in this way can be used as an approximate indication of the direction of travel of the disturbances. If the disturbance appears on the ordinary ray first then the disturbance must be travelling towards the equator.

The equivalent frequency examination may be applied at frequencies other than the critical if the indications are clear enough on both ray traces; for example, if two frequencies are selected differing by $f_x^0 - f_0^0$ and we can observe the time difference of occurrence of disturbances on the x-ray of the higher frequency and the o-ray of the lower frequency, which would be reflected from the same height, this will be the time of travel between the two reflection points. If, then, the horizontal speed and direction are known from three-point observations the separation can be determined. Furthermore, if the time difference of occurrence on the ordinary ray at these two frequencies is observable, the height difference of these two reflection points can be determined by applying analyses derived by Manning (1947) or Kelso (1952) to the undisturbed $h'f$ record, and so the vertical velocity may be deduced. From the vertical and the horizontal velocities the slope of the front can then be calculated. These time differences can be determined more accurately by $h't$ recording on two appropriate fixed frequencies. This method is being used to determine the slope of fronts of disturbances.

III. Fixed Frequency Records

In the study of travelling disturbances the virtual heights of echoes from three transmitters spaced some 40 km apart and operating on the same frequency are all recorded on a single film. It is quite often found that complexities caused by travelling disturbances appear on both ordinary and extraordinary traces for all three transmissions. On most occasions the disturbance occurs first on the ordinary ray, and for the purpose of the present paper the time difference between the ordinary and the extraordinary ray disturbance in this case is considered as positive. Two examples of these disturbances are shown in Fig. 4. In Figure 4 (*a*) there is a positive time difference of 4 min between appearances on the *o*- and *x*-rays, but in Figure 4 (*b*) appearances are almost simultaneous.

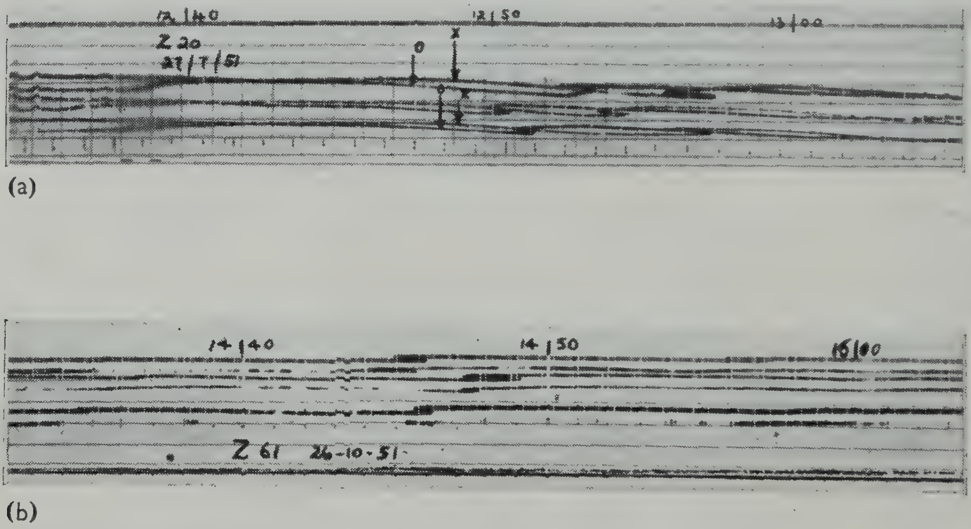


Fig. 4.—*Z* type disturbances showing time differences between *o* and *x* disturbance effects.

Referring to Figure 2, at any given frequency the *o*- and the *x*-rays deviate in the Earth's magnetic meridian plane and are reflected at two points A and B in the F_2 region. It will be seen, therefore, that a horizontally travelling disturbance will in general reach points A and B at different times.

If the disturbance has a vertical front and is travelling in the direction of the Earth's magnetic field, the horizontal separation x is equal to $A'B$, which is equal to Vt_1, where V is the velocity of the disturbance and t_1 is the time difference of occurrence. If the disturbance is travelling in a direction at an angle χ to the Earth's magnetic field, then

$$x = Vt_1/\cos \chi.$$

If the disturbance has a vertical front and is travelling in a direction normal to the magnetic meridian plane, then it should arrive at points A and B simultaneously and t should be zero.

It has been found in practice, however, that disturbances travelling normally to the magnetic meridian plane always show a positive time difference between o- and x-ray disturbance effects. This must be due to the disturbance front possessing a slope γ to the horizontal as shown in Figure 2, and the vertical separation may be expressed as $Vt_2/\cot\gamma$, where t_2 is the time difference contributed by the sloping front of the disturbance. It follows therefore that:

$$x\cos\chi+z\cot\gamma=V(t_1+t_2),$$

or

$$x\cos\chi+z\cot\gamma=Vt,$$

where t is the total time difference.

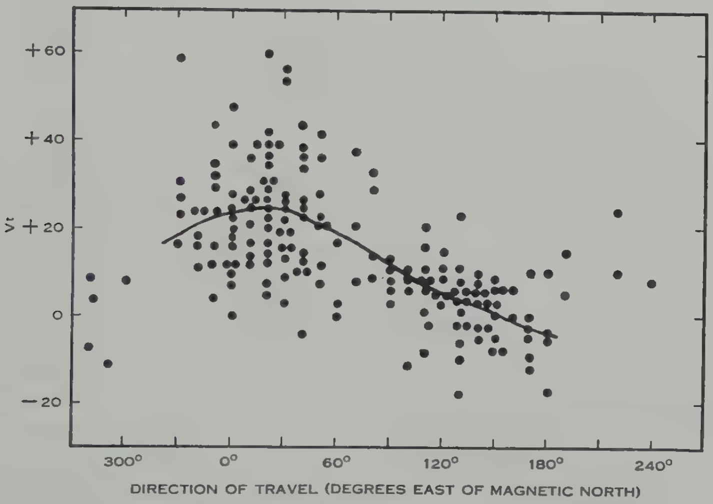


Fig. 5.—Plot of Vt *against* direction of travel for complexities on o- and x-rays. April 1, 1950 to March 31, 1951.

Figure 5 shows a plot of Vt for complexities on o- and x-rays for a period of 12 months from April 1, 1950 to March 31, 1951. These are plotted against directions E. of the magnetic meridian.

In each case considered, the disturbance indication is a complexity (of the loop, Y, or Z type; Munro 1953). The time of occurrence of this disturbance can be recorded to the nearest $\frac{1}{4}$ min and the disturbances are chosen so that their directions of travel as recorded for the o and x traces do not differ by more than 20°. Moreover, the disturbances must be present on both traces for all three of the spaced transmitters to enable this discrimination to be applied.

These criteria ensure a reasonable uniformity in type of disturbances considered. They must have sensibly linear fronts of dimensions greater than the spacing of stations, and be of magnitude great enough to give clear discontinuities, but not great enough to cause serious distortion of the normal ionization gradient. This screening appreciably reduced the scatter of points.

Observational accuracy has been examined and is considered to be sufficient to ensure significance of results. The consistency of results confirms this.

The following features are immediately apparent. There is a definite mean trend from high positive values of Vt at 0° to mainly negative values at 180°. Between 0 and 90°, there are many high points with values from 10 to 40, a few points with values from 0 to 10, and very few negative points ; whereas, between 90 and 200° there are very few positive points greater than 10 and many negative ones. This, therefore, is definite confirmation of the general correctness of concepts previously expressed. The scatter of points may be partly due to variations in slope of front and observational limitations as well as to actual differences in separation.

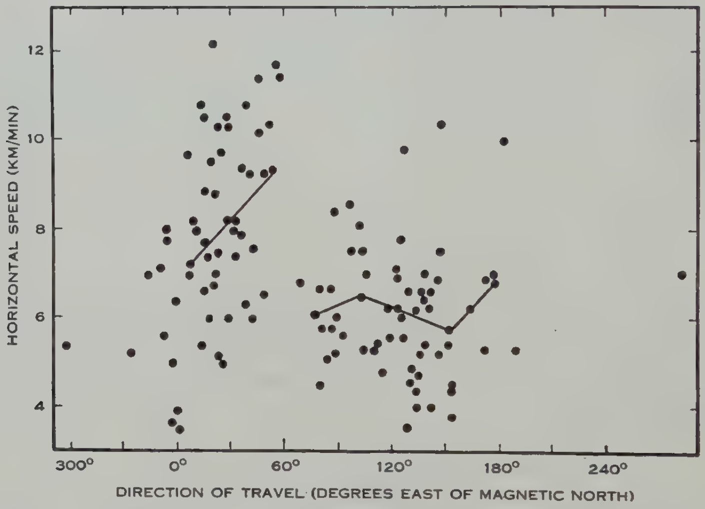


Fig. 6.—Variation of speed with direction of travel for disturbances in Figure 5.

Variations in horizontal speed have been examined and while there is no correlation apparent between speed and time difference on *o*- and *x*-rays for a given direction, there is a definite variation of speed with direction. Figure 6 is a plot of speed against direction for the same disturbances as in Figure 5. It will be seen that there is a large change of speed between 60 and 70°. This corresponds closely with the seasonal change in speed, as directions 0 to 50° are recorded mainly in winter when speeds are highest, and other directions at the equinoxes and in summer when speeds are relatively low.

The results for the succeeding year April 1, 1951 to March 31, 1952 were similar and are shown in Figures 7 (*a*) and 7 (*b*). The velocity change for this year is not as marked. The similarity of result for the two years firmly establishes the fundamental features confirming general theory. We may therefore make certain deductions on the basis of that theory.

Since the disturbances travelling in directions close to 90° will be moving normally to the Earth's magnetic meridian plane, values of Vt other than zero must be due entirely to the vertical separation of the o and the x reflection points, and if Vt is positive there must be an apparent component of progression vertically downward, which on our assumption would imply a forward tilt of the disturbance.

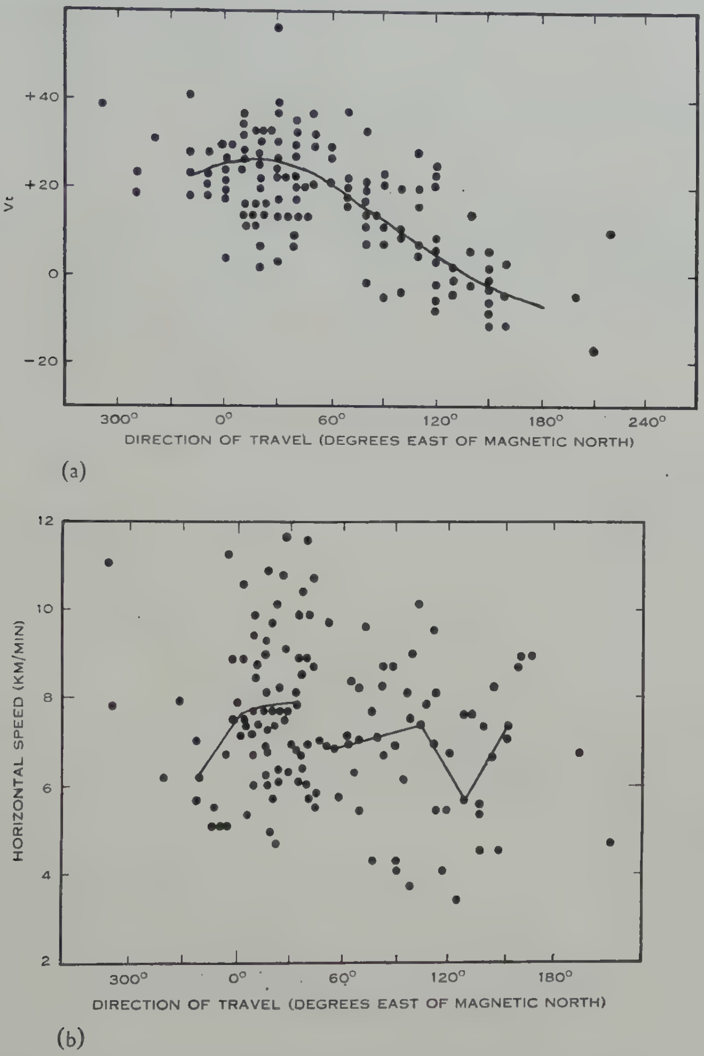


Fig. 7 (a).—Plot of Vt *against* direction of travel for complexities on o- and x-rays. April 1, 1951 to March 31, 1952.

Fig. 7 (b).—Variation of speed with direction of travel for disturbances in Figure 7 (a).

It will be seen in Figure 5, that all values of Vt between 80 and 100° are positive. This supplies strong evidence that there is a forward slope of front in all cases. This is supported also by the fact that the negative values between 100 and 150° are much smaller than the positive values between 0 and 100°.

To derive a mean curve, all the values of Vt for disturbances occurring from April 1, 1948 to March 31, 1954 have been averaged and plotted in Figure 8 and the best curve drawn through them. Where the data are most reliable this curve suggests a cosine curve and confirms the validity of the expression

$$x \cos \chi + z \cot \gamma = Vt. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

This equation is of the form

$$a \cos \chi + b = c,$$

giving $x=14$ and $z \cot \gamma = 9$ for the curve of Figure 8. In Figure 10 Kirkpatrick shows that in Sydney $x/z=0{\cdot}8$, therefore, $z=17{\cdot}5$. At $\chi=0°$, $\gamma=65°$, which means that when a disturbance is travelling in the direction of the Earth's magnetic meridian its front is tilted at an angle of 65° to the horizontal. At $\chi=90°$, $\gamma=54°$, so that the disturbance still has a tilted front when travelling at right-angles to the Earth's magnetic meridian. γ has a minimum value of 51° when $\chi=120°$. In Sydney, the magnetic dip angle is $63{\cdot}9°$ so that a disturbance travelling along the Earth's magnetic meridian travels with its front approximately parallel to the field.

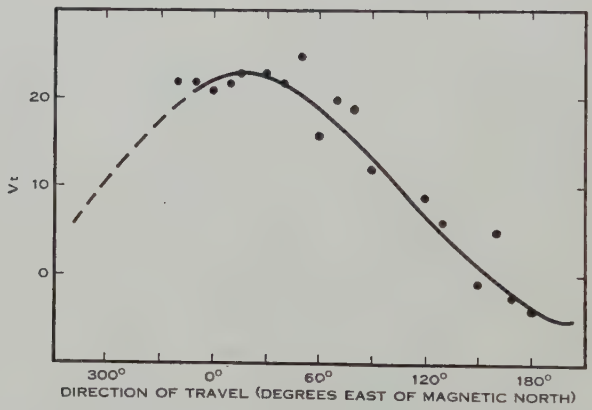


Fig. 8.—Variation of mean Vt with direction of travel. April 1, 1948 to March 31, 1954.

It will be noticed in Figure 8 that the curve is not symmetrical about 90°. The value of Vt at 90° is 13 and at 0° is 23, a difference of 10, whereas, the value at 180° is −5, a difference of 18 from the value at 90°. This asymmetry could result from the mean slope of front, γ, being greater in winter than in summer. This is supported by other evidence which will be discussed in another paper.

Since, as mentioned above, directions vary with the seasons, and the sign of the time difference alone is an indication of general direction, it is of interest to consider these data on a seasonal basis. In Figure 9, all the recorded values of t are plotted for each month. The dots are the data used in Figure 5 selected on a directional consistency basis, and the crosses are the rejected values. It will be seen that their inclusion increases the scatter but has little effect on mean values. The scarcity of negative points in the winter months and their frequency in summer months is very marked.

It will be seen therefore that a seasonal plot of t for disturbances on o and x traces at a single station on a fixed frequency can give quite definite information as to the seasonal trend of movement.

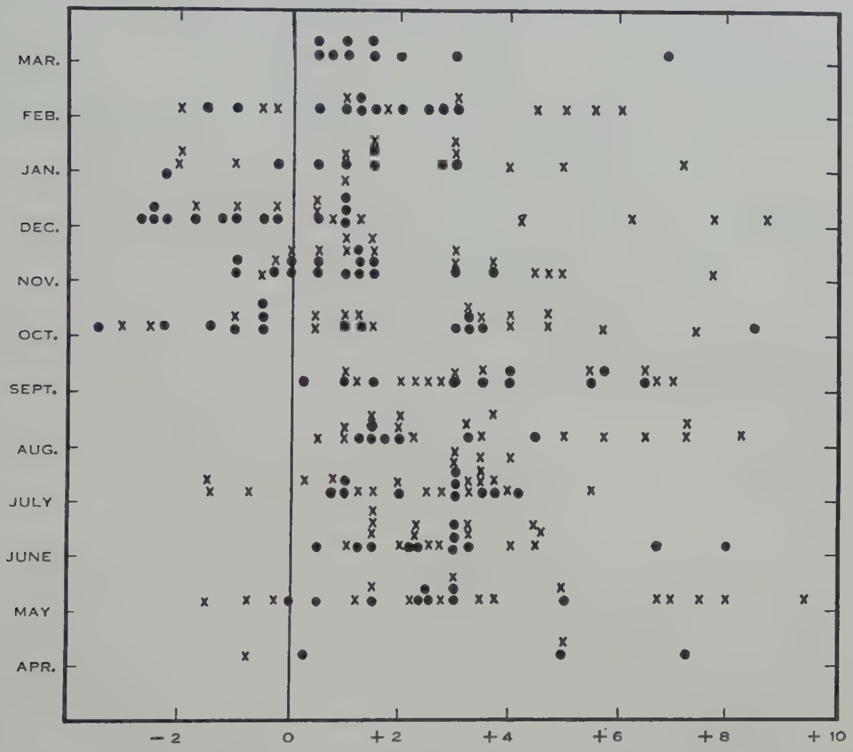


Fig. 9.—Seasonal variation of time-differences on o- and x-rays 1950–51. Dots are selected cases used in Figure 5 and crosses are cases considered insufficiently reliable for use in Figure 5.

IV. Interpretation of Records

It follows from the above results that certain features of disturbance irregularities on ionospheric records will vary according to the direction of travel of the disturbances. When the disturbance is travelling in the direction of the Earth's magnetic field its effect, usually apparent as an increase in virtual height, will appear first on the o-ray and a few minutes later on the x-ray. On a fixed frequency record, this generally produces crossing over of the two traces and the formation of loops as shown in Figure 4 (*a*).

If the disturbance is travelling in a direction normal to the field, the two rays will be affected almost simultaneously and the traces will move up and down together. This may be seen in Figure 4 (*b*).

On variable frequency ($h'f$) records the effects are more complex, but the general feature is a perturbation appearing first at the top of the trace for one ray and progressing downward with some progressive change of form. Similar changes will appear on the other trace, generally with a time difference at any given height. As on $h't$ records, when the disturbance is travelling along the field and the time difference is great, there will be considerable alteration in the relative position of the two traces; for example, under steady conditions the F_1 retardation peak on the x-ray usually brings this trace above the o trace in virtual height for part of the frequency range. A disturbance may temporarily reduce this peak and increase the height of the o trace so that the two traces are separated for their full length. However, if the disturbance is travelling across the field any time difference is small and an alteration in the relative position of the two traces may not be evident.

It should be noted that a travelling disturbance can cause a change in penetration frequency and, if this occurs at different times on the two rays, there will be fluctuations in the difference in the o and x critical frequencies at a given time. Therefore, in making deductions (e.g. the value of H) from this difference, care must be taken to avoid, or allow for, disturbance effects. Only if the disturbance is crossing the field will the perturbations affect both traces at the same height at the same time.

V. Conclusions

As the information from the two types of recording are supplementary and in agreement we may summarize the conclusions as follows:

Both fixed frequency and variable frequency observations confirm the theory of divergence of the o- and the x-rays in the F region at Sydney, with an actual separation of the order of 30 km near the maximum of ionization.

For a fixed frequency of 5·8 Mc/s over the period of observation March 1950 to April 1954, the average horizontal separation of o and x reflection points at Sydney is 14 km while the average vertical separation is 17·5 km.

The observations also indicate that a travelling disturbance always has an apparent vertical component of progression which is here assumed to be the result of a forward tilt in the wave front of the disturbance.

When the disturbance is travelling in the direction of the Earth's magnetic field, the tilt $\gamma=65°$, so that the front of the disturbance in Sydney is almost parallel to the Earth's magnetic field (dip angle in Sydney is 63·9°). This may be of significance in determining the nature of travelling disturbances. Furthermore, if the disturbance is travelling in a direction normal to the Earth's magnetic field, there is an appreciable tilt of its front. A "cellular wave" theory of disturbances advanced by Martyn (1950) suggests that such disturbances should have no tilt so that some modification seems to be required.

With a knowledge of disturbance effects on records, it is also possible from a set of $h'f$ records to deduce an approximate direction of travel of the disturbances. For this purpose, it is desirable to have records at intervals of 1 to 2 min. This method is being investigated for extending the knowledge of disturbance movements over a wide area.

Conditions in this paper apply particularly to the geomagnetic latitude of Sydney. Corresponding effects to be expected in other places may be deduced from a knowledge of the appropriate magnetic constants and the use of Figure 10 to determine the ratio of horizontal to vertical separation.

From the curves of Figure 10, the horizontal separation per kilometre of vertical penetration of the layer for a frequency of 5·8 Mc/s at Sydney is approximately 0·2 km. Since the mean observed separation is 14 km this implies that the divergence commences some 70 km below the *o*-ray reflection height thus giving an indication of the thickness of the ionized region during the day-time.

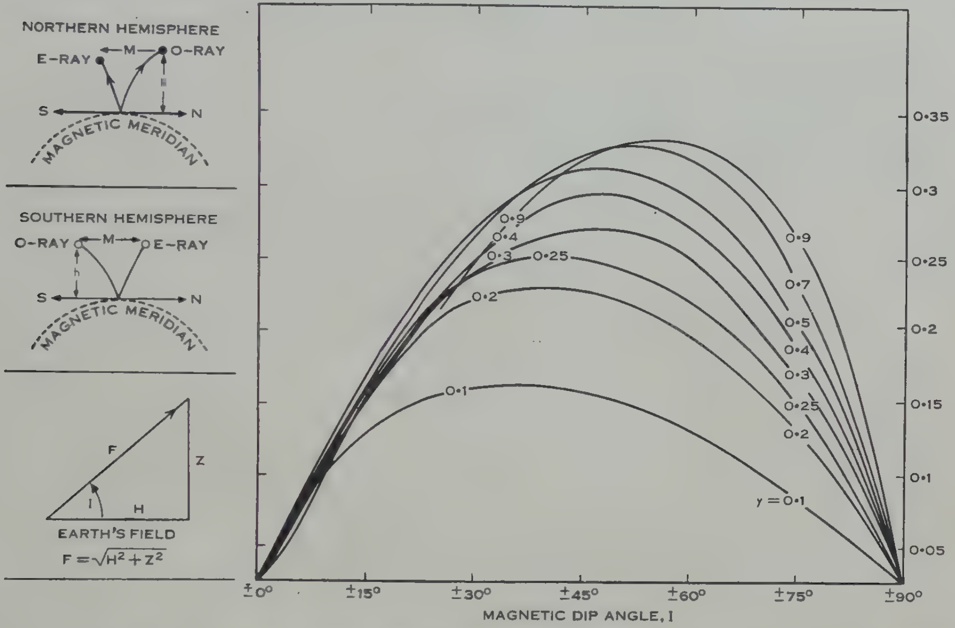


Fig. 10.—Chart giving ratio of horizontal to vertical separation for various values of magnetic dip angle (Kirkpatrick).

m=total horizontal separation in km per km of vertical penetration by the *o*-ray.
h=height of *o*-ray reflection point above base of layer.
$(1-y)h$=height of *x*-ray reflection point above base of layer.
$M=mh$=total (horizontal) separation.
yh=vertical separation.
f_H=gyrofrequency=$2\cdot 82F$ (Mc/s).
F=total magnetic field strength (gauss or Γ).
f=wave frequency (Mc/s).
$y=f_H/f$.

At Sydney, $F=0\cdot 511$, $f_H=1\cdot 43$ Mc/s.

VI. ACKNOWLEDGMENTS

The detailed reduction of records and preparation of material by the analysis staff of the Radio Research Section, Sydney, are gratefully acknowledged, and the authors are particularly indebted to Mr. C. B. Kirkpatrick of the Mathematics

Department, New South Wales University of Technology, for helpful discussions on theoretical considerations. They wish to express their appreciation to the University of Sydney and in particular to Professor D. M. Myers for provision of facilities in the Electrical Engineering Department.

VII. REFERENCES

BOOKER, H. G. (1938).—*Phil. Trans.* A **237** : 425.
HEISLER, L. H. (1955).—*Aust. J. Appl. Sci.* **6** : 1.
KELSO, J. M. (1952).—*J. Geophys. Res.* **57** : 357.
MANNING, L. A. (1947).—*Proc. Inst. Radio Engrs. N.Y.* **35** : 1203.
MILLINGTON, G. (1949).—*Nature* **163** : 213.
MUNRO, G. H. (1949).—*Nature* **163** : 812.
MUNRO, G. H. (1950).—*Proc. Roy. Soc.* A **202** : 208.
MUNRO, G. H. (1953).—*Proc. Roy. Soc.* A **219** : 447.
MUNRO, G. H., and HEISLER, L. H. (1956).—*Aust. J. Phys.* **9** : 343.

THE THEORY OF STRONG KATABATIC WINDS

By F. K. Ball*

[*Manuscript received February* 24, 1956]

Summary

In regions where strong katabatic winds occur, notably Commonwealth Bay, Antarctica, there is often a sharp boundary between the strong winds and the comparatively calm conditions at sea. The seaward movement of this boundary across a coastal station is accompanied by a sudden onset of strong winds and the landward movement by their sudden cessation. The boundary has all the characteristics of a pressure jump line accompanying a sudden change in depth of the cold air. Furthermore, it is to be expected from theoretical considerations that a pressure jump will occur near the coast. A simple theory is developed from which various characteristics of the flow can be predicted, including the intensity of the jump and the deviation of the wind from the lines of greatest slope caused by the Earth's rotation. These predictions agree moderately well with observations.

I. Introduction

According to the meteorological glossary katabatic winds blow down slopes that are cooled by radiation, the direction of flow being controlled almost entirely by orographic features. Winds of this type are best developed on the large ice caps of Greenland and Antarctica where the surface winds, apart from a slight deviation attributed to the Earth's rotation, almost always blow down slope, regardless of the direction. Near the foot of the ice slope these winds may be extremely violent and exhibit certain peculiarities which have excited comment from various authors, notably Mawson (1915), Madigan (1929), Mirrlees (1934), Prudhomme and Boujon (1952), and Boujon (1954). The peculiar features are perhaps best shown by the winds in the neighbourhood of Commonwealth Bay near the border of King George V Land and Adélie Land, Antarctica, though they occur at other places along the Antarctic coast and elsewhere. Only the principal features will be mentioned here; the reader is referred to the above-mentioned papers for a more detailed description.

The winds are extremely strong, persistent, and steady both in speed and direction. However, the striking feature is the occurrence of curious lulls during which the wind is light and variable. At the commencement of a lull the wind drops with great suddenness and almost simultaneously the pressure rises abruptly by 2–3 mb. Similarly, at the end of a lull the wind rises suddenly and the pressure falls abruptly. During the period of the lull strong winds may be both audible and visible higher up the slope and sometimes a strong airstream, rendered visible by drift snow, is seen overriding the calmer air beneath. A rotating roll cloud may also be present. These observations suggest the existence of a sharp boundary between the strong katabatic winds and the comparatively calm air over the sea. It will be shown subsequently that a sharp boundary is to be expected on theoretical grounds.

---

* Division of Meteorological Physics, C.S.I.R.O., Melbourne.

The main purpose here is the theoretical investigation of strong katabatic winds with particular reference to the retarding stage near the foot of the slope where the peculiar features have been observed. High up on the slope the air must pass through an accelerating stage. Probably before reaching the foot an approximate equilibrium is set up between the frictional resistance and the katabatic force, i.e. the wind approaching the coast is antitriptic. Near the coast the inertia of the wind plays a dominant part so that any satisfactory theoretical description of the retarding stage must take account of pressure forces, friction, and inertia. The problem will be simplified here by assuming the steady state, taking the upwind supply of cold air as given and supposing the potential temperature of the katabatic flow to be constant. The Earth's rotation can also have an appreciable effect on the flow but this will be neglected in the initial formulation of the theory, discussion being postponed to Section V.

II. Basic Equations

Pilot balloon observations indicate that the katabatic winds in the neighbourhood of Commonwealth Bay can be considered as a relatively thin layer of cold air flowing down the slope under the influence of gravity. For simplicity it will be supposed that the cold air has a definite depth, i.e. it is surmounted by a sudden inversion, and that the velocity and potential temperature of the cold air are independent of height. If there is no mixing across the inversion and cross-slope variations are neglected then the equation of continuity in the steady state takes the form

$$\frac{\mathrm{d}}{\mathrm{d}x}(hu)=0, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where h is the height of the inversion, u is the velocity, and x is the horizontal distance measured downstream from an arbitrary fixed reference point. If in addition to the above assumptions vertical accelerations can be neglected, then the equation of motion can be written

$$\frac{\mathrm{d}}{\mathrm{d}x}(hu^2)=\frac{h\theta' g\alpha}{\theta}-\frac{h\theta' g}{\theta}\frac{\mathrm{d}h}{\mathrm{d}x}-\frac{h}{\rho}\frac{\mathrm{d}p}{\mathrm{d}x}-ku^2, \quad \ldots\ldots\ldots \quad (2)$$

where θ=potential temperature,

θ'=potential temperature deficit of the cold air,

α=angle of inclination of the ice slope (supposed small),

$\frac{\mathrm{d}p}{\mathrm{d}x}$=pressure gradient in the air above the inversion, i.e. the superimposed pressure gradient,

k=a dimensionless constant.

The last term on the right-hand side of equation (2) represents the friction force, supposed proportional to the square of the velocity, this being the generally accepted law for surface friction with moderate to strong winds. The other three terms arise from the horizontal pressure gradient on the assumption that the vertical distribution of pressure is hydrostatic. They will be referred to as the katabatic force, dependent on the slope, the gradient force, dependent on the gradient of h, and the force due to the superimposed pressure gradient. The sum of the first two terms vanishes when $\mathrm{d}h/\mathrm{d}x=\alpha$, i.e. when the inversion surface is horizontal. The direct effect of the superimposed pressure gradient in the neighbourhood of the foot of the slope is usually an order of magnitude smaller than the other terms and will henceforth be neglected. It may be taken into account by modifying α appropriately.

Equation (1) integrates immediately giving

$$hu=Q, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

where Q is a constant depending on the rate of supply of cold air from the ice plateau. Substitution in equation (2) gives an equation in h alone

$$\frac{\mathrm{d}h}{\mathrm{d}x}\left(h^3-\frac{\theta Q^2}{\theta' g}\right)=\alpha\left(h^3-\frac{k}{\alpha}\,\frac{\theta Q^2}{\theta' g}\right). \quad \ldots\ldots\ldots\ldots \quad (4)$$

This equation was first formulated and solved by Bresse in connexion with the flow of water in open channels. It has been considered in detail by hydraulic engineers who refer to its solutions as "backwater curves". Much of what follows in this section and Sections III and IV is based on standard hydraulic theory.

The critical depth h_c is defined by

$$h_c^3=\theta Q^2/\theta' g. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

If h equals h_c then the second factor on the left-hand side of equation (4) vanishes, therefore either $\mathrm{d}h/\mathrm{d}x$ is large and the equation is no longer valid since vertical accelerations cannot then be neglected, or the right-hand side vanishes, in which case α equals k and the slope is said to be critical. The flow will be termed tranquil or shooting according as h is greater than or less than h_c. In a region of flow throughout which equation (4) is valid the change from tranquil to shooting flow or vice versa can only occur where the slope is critical. In hydraulic applications it is found that equation (4) gives a sufficiently accurate description of the flow except in regions, covering only a very small range in x, where the surface changes abruptly from one level to another. These regions are known as jumps, and it will be shown in Section III that upstream of a stationary jump the flow is shooting and downstream it is tranquil. It can be seen therefore that a jump provides a means whereby the flow can change from shooting to tranquil at a point where the slope is not necessarily critical.

When the air has ceased to accelerate, $\mathrm{d}h/\mathrm{d}x$ is zero and the depth of the flow, called the normal depth h_n, is given by

$$h_n^3=k\theta Q^2/\alpha\theta' g. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

The uniform flow is shooting if $h_n<h_c$, i.e. if the slope is greater than the critical.

The preceding results can be expressed conveniently in terms of the Froude number, a dimensionless number defined as

$$F=\theta Q^2/\theta' g h^3. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

When the depth of the flow is critical F is unity, and the flow is shooting or tranquil according as F is greater or less than unity. The Froude number of the uniform flow is α/k and thus depends only on the steepness and roughness of the slope and not on the temperature or rate of supply of cold air. Various physical interpretations of the Froude number can be made. It can be regarded as the ratio of the inertia force, ρu^2, to the modified gravitational force, $\rho\theta' gh/\theta$, implying that the inertia of the flow dominates over gravity in shooting flow and vice versa in tranquil flow. It may also be regarded as the square of the ratio between the velocity, u, and the velocity of propagation of small amplitude gravity waves, $(\theta' gh/\theta)^{\frac{1}{2}}$, whence it follows that these waves cannot travel upstream in shooting flow. Finally it may be regarded as the reciprocal of a finite difference form of the Richardson number, i.e. with temperature gradient replaced by θ'/h and velocity gradient by u/h.

The stability of uniform flow has been investigated, among others, by Jeffreys (1925) with reference to flow of water in channels and by Defant (1933) with reference to air flow. It is found by simple perturbation theory that the flow is stable provided $F<4$. If $F>4$ then roll waves should develop. A description of this type of wave is given in Cornish (1934). Observations show that the roll waves do not develop except on slopes considerably steeper than indicated by the theory, partly because a very great length of slope is required for the waves to build up on slopes near the minimum.

Equation (**2**) expresses the momentum balance of the cold air ; the energy balance equation can be found by multiplying through by u and using equation (**3**) giving

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left[\tfrac{1}{2}hu^3+\tfrac{1}{2}\frac{\theta' g}{\theta}h^2u+\frac{\theta' ghu}{\theta}(z+\tfrac{1}{2}h)\right]=ku^3, \quad \ldots\ldots\ldots\ldots \quad (\mathbf{8})$$

where $\mathrm{d}z/\mathrm{d}x=-\alpha$, whence z is the height of the surface above some arbitrary fixed reference level. The three terms in the square brackets on the left are respectively, the rate of transport of kinetic energy, the rate of working of the gradient force, and the rate of transport of potential energy. The total rate of decrease of energy is given by the rate of working of the friction force. In this case therefore the energy equation is implied by continuity and the equation of motion and no additional information can be obtained from it. It will be shown in Section III that this is not the case at a jump or sudden discontinuity in the flow.

III. CONDITIONS AT A JUMP

Conditions at a jump may be investigated by applying conservation of mass and momentum to the flow in the neighbourhood of the jump. The change in depth of the cold air, though sudden, will occur over a finite distance which will be termed the jump width. It will be assumed that the changes in energy and momentum caused by the sudden change in depth outweigh the changes caused by surface friction and slope within the jump width. If the suffixes 1 and 2 are used to denote upstream and downstream values respectively, then, from continuity at a stationary jump aligned perpendicularly to the direction of flow,

$$h_1u_1=h_2u_2=Q, \quad \ldots\ldots\ldots\ldots \quad (9)$$

and, by equating the force to the rate of change of momentum,

$$h_1u_1^2+\tfrac{1}{2}h_1^2\,\frac{\theta' g}{\theta}=h_2u_2^2+\tfrac{1}{2}h_2^2\,\frac{\theta' g}{\theta}=M. \quad \ldots\ldots\ldots\ldots \quad (10)$$

Q is a known constant for the whole flow, both upstream and downstream of the jump, whereas the value of M is changed by slope and friction effects and is not known. The above relations therefore constitute three independent equations which enable any two of the four variables h_1, h_2, u_1, and u_2 to be eliminated.

Elimination of u_1 and u_2 leads to the symmetrical relation

$$h_c^3=h_1h_2(h_1+h_2)/2. \quad \ldots\ldots\ldots\ldots \quad (11)$$

Therefore if $h_1 \lessgtr h_c$ then $h_2 \gtrless h_c$ and from these considerations a stationary jump can only occur where the flow changes from shooting to tranquil or vice versa, i.e. the flow must jump through the critical depth. For purposes of calculating the change in depth, given either upstream or downstream conditions, equation (**11**) may be solved for either h_1 or h_2 giving

$$\left.\begin{aligned} h_2&=\tfrac{1}{2}h_1[(1+8F_1)^{\frac{1}{2}}-1], \\ h_1&=\tfrac{1}{2}h_2[(1+8F_2)^{\frac{1}{2}}-1]. \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (12)$$

For a given value of h_c, i.e. given values of Q and θ'/θ, then from equation (**12**) there corresponds to each depth h_1 a depth h_2 which will be denoted by a circumflex, thus $h_2=\hat{h}_1$ or $h_1=\hat{h}_2$.

The rate of change of mean flow energy per unit length of jump is given by

$$\Delta E=\rho\left[\tfrac{1}{2}h_2u_2^3-\tfrac{1}{2}h_1u_1^3+\frac{\theta' g}{\theta}(h_2^2u_2-h_1^2u_1)\right]=\frac{\rho Q^3(h_2-h_1)^3}{4h_1h_2h_c^3}. \quad \ldots \quad (13)$$

Since energy cannot be created at a jump, ΔE must be negative, therefore $h_2>h_1$ and *a stationary jump can only occur as a transition from shooting to tranquil flow.* The mean flow energy which is lost at the jump either appears as turbulent kinetic energy or is radiated away by a stationary wave system situated downstream of the jump. For small hydraulic jumps the stationary wave system accounts for most of the energy whereas the energy appears as turbulence in intense jumps. These features of hydraulic jumps have been discussed by Benjamin and Lighthill (1954) and Binnie and Orkney (1955).

The equations for a stationary jump may readily be extended to apply to a moving jump, provided Q is replaced by Q_j the rate of flow through the jump. If the jump is moving downstream with a velocity c, then

$$Q_j=h_1(u_1-c)=h_2(u_2-c). \quad \ldots\ldots\ldots\ldots \quad (14)$$

The upstream Froude number relative to the jump is F_{1j} say, where

$$F_{1j}=F_1(u_1-c/u_1)^2. \qquad (15)$$

In order to calculate the change in height at a moving jump, F_1 must be replaced by F_{1j} in equation (**12**). It follows that, for given values of h_1 and F_1, a jump moving upstream is more intense than a stationary one and a jump moving downstream is less intense. Similar considerations apply if h_2 and F_2 are given. In general therefore, a jump moving upstream tends to be stronger and a jump moving downstream tends to be weaker than a stationary one.

The normal jump equations can also readily be extended to apply to a jump inclined to the direction of flow. Equations (**9**)–(**13**) are still valid provided u_1 and u_2 are taken as the components perpendicular to the jump line. The wind component parallel to the jump line is unchanged, whereas, at an intense jump, the component perpendicular to the jump is reduced almost to zero. In these conditions, even if the incoming flow is only inclined at a small angle to the normal, the outflowing air moves almost parallel to the jump line. The relationship between the inclinations to the normal, β_1 and β_2, before and after the jump can be deduced immediately from equation (**12**) by the substitution $h=Q/u=Q\tan\beta/v$, giving

$$\tan\beta_2=\tfrac{1}{2}\tan\beta_1\{(1+8F_1)^{\frac{1}{2}}-1\}. \qquad (16)$$

IV. The General Solution

If α is not zero the general solution of equation (**4**) with α and k constant is

$$\alpha x/h_n=h/h_n+(1-F_n)\Phi(h/h_n), \qquad (17)$$

where Φ is Bresse's Backwater Function given by

$$\Phi(s)=\frac{1}{6}\log\left\{\frac{s^2+s+1}{(s-1)^2}\right\}+\frac{1}{\sqrt{3}}\arctan\left\{\frac{s\sqrt{3}}{1-s}\right\}, \qquad (18)$$

and F_n is the Froude number for uniform flow. This function is tabulated in Woodward and Posey (1941), Table 601. When α is zero the solution takes the particularly simple form

$$kx/h_c=h/h_c-\tfrac{1}{4}(h/h_c)^4. \qquad (19)$$

The origin of x has been taken, for convenience, where h vanishes.

In a region throughout which equation (**4**) is valid, two boundary conditions are required to determine the solution completely, e.g. the value of Q and the depth h_0 at some known value of x, x_0 say. In practice for given Q this value of x is upstream in shooting flow and downstream in tranquil flow, because the behaviour of shooting flow is determined physically by upstream conditions and that of tranquil flow by downstream conditions. This is indicated when, for instance, an attempt is made to determine the depth some distance upstream in shooting flow where downstream conditions are given. It is found that very small differences in the given depth downstream can produce extremely large differences in the predicted depth upstream.

There are 12 different types of solution which can occur, only four of which will be considered here. These are:

(i) If $h_0 > h_n > h_c$, then $\mathrm{d}h/\mathrm{d}x$ is positive and tends towards α for large x, so that the inversion becomes horizontal downstream. For large negative x, h tends to h_n. The flow is tranquil.

(ii) If $h_0 > h_c > h_n$, then $\mathrm{d}h/\mathrm{d}x$ is positive and, as previously, the inversion surface becomes horizontal downstream. The depth decreases upstream until it approaches the critical depth where the assumed conditions no longer apply. The flow is also tranquil in this case.

(iii) If $h_0 = h_n$, then $\mathrm{d}h/\mathrm{d}x$ is zero and the flow is uniform. The flow may be either tranquil or shooting according to whether $h_n \gtrless h_c$.

(iv) If $h_c > h_0$ and $\alpha = 0$, whence h_n is undefined, then once more $\mathrm{d}h/\mathrm{d}x$ is positive and the depth increases downstream until it approaches the critical depth where the assumed conditions no longer apply. The flow is shooting.

Solutions (i), (ii), and (iv) are shown in Figure 1.

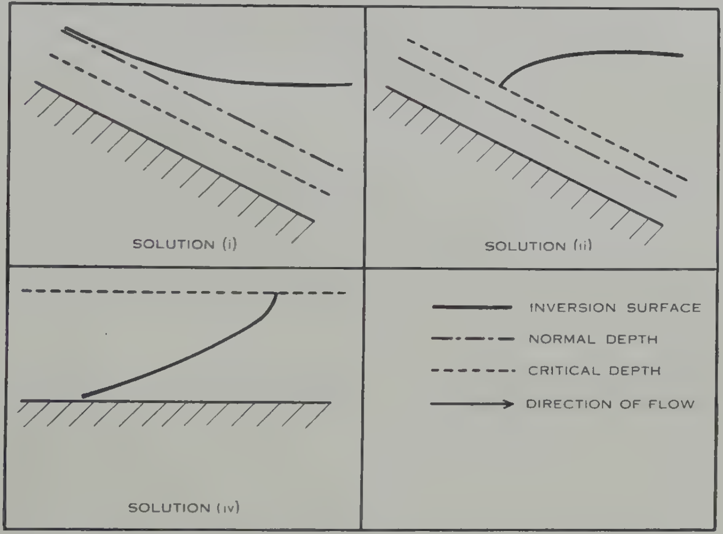


Fig. 1.—Backwater curves involving retardation

We are now in a position to determine the behaviour of the katabatic flow in the retardation stage near the coast. The rate of supply of cold air, Q, is taken as given and the flow is assumed to be uniform before it is retarded by coastal influences. The characteristic depth, H say, of the cold air over the sea is also taken as given. This depth cannot be predicted on the present theory since many of the factors which control the movement of the air over the sea have been neglected, e.g. the warming of the air by contact with the sea, friction from upper winds, and the effect of the Earth's rotation. These effects and others too will become important once the air has been retarded. It will also be assumed that $H > h_c$ since there appear, in general, to be no forces sufficient to maintain shooting flow at sea.

In the preceding paragraph three boundary conditions have been specified, the rate of flow Q, the upstream depth h_n, and the downstream depth H. If equation (4) were valid throughout the whole region one of these conditions would be unnecessary. However, the occurrence of a jump divides the flow into two separate regimes each of which, assuming Q known, requires one boundary condition. Furthermore, an additional condition is required to fix the position of the jump. This is provided by the jump equation (**12**) which relates the depths at the common boundary of the two regimes.

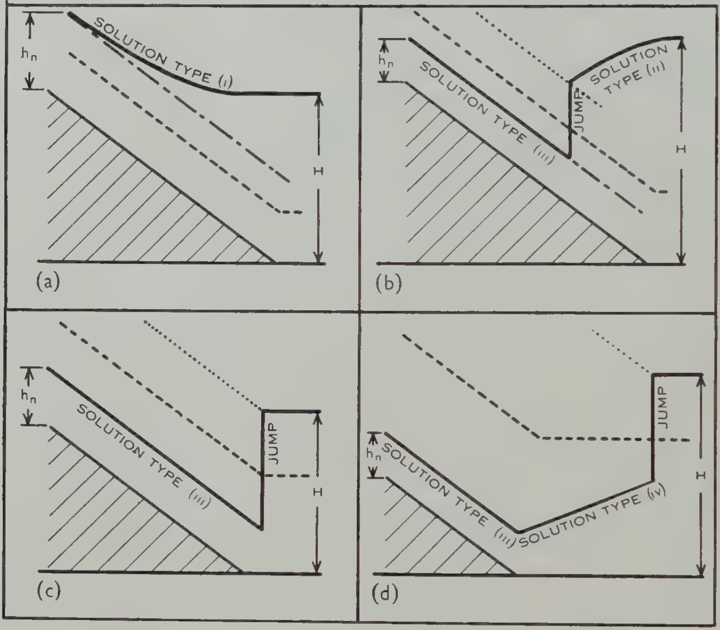


Fig. 2.—Flow types which can occur near the foot of a slope.
——— Inversion surface h. — — — — — Normal depth.
- - - - - Critical depth h_c. ······· $\hat{h}$.
——→ Direction of flow.

The behaviour of the katabatic wind depends firstly on whether the uniform flow is shooting or tranquil. If the flow is tranquil, then no jump can occur and the air is retarded before reaching the coast, the depth of cold air being given by a solution of type (i). For given Q the solution is completely specified if H is known. The upstream condition $h=h_n$ is automatically satisfied for all solutions of type (i). A coastal station would in this case be in a region of light winds. In the neighbourhood of Commonwealth Bay this kind of flow does not usually occur because, as will be shown subsequently, h_n is normally less than h_c.

When the uniform flow is shooting a jump will occur near the coast. If the jump is situated inland then the flow between jump and coast will be type (ii) and the flow upstream of the jump will be normal, i.e. type (iii). A coastal station will once more be in a region of light winds since the air is retarded before reaching the coast. It is in this case that strong winds may be both audible and visible higher up the slope, as noted by various observers in both the Antarctic and Greenland. If the jump is situated out to sea then the flow between jump and coast will be of type (iv), the flow on the slope being uniform as before. It is in this case and this case only that strong winds occur at the coast. The flow types are shown in Figure 2.

When H equals $\hat{h}_n$ then the jump will be situated at the coast. Furthermore, it will be inland or at sea according as H is greater than or less than $\hat{h}_n$. The approximate position of the jump when it is near the coast can easily be calculated. Consider first the case $H>\hat{h}_n$. If H is the depth at the coast then it changes by an amount Δh at a distance inland from the coast given by

$$x=-\left(\frac{H^3-h_c^3}{H^3-h_n^3}\right)\frac{\Delta h}{\alpha}, \qquad \text{(20)}$$

provided Δh is small. This is merely a restatement of equation (**4**) in a finite difference form. Immediately downstream of the jump h equals $\hat{h}_n$, so the distance of the jump from the coast is determined by putting Δh equal to $H-\hat{h}_n$ in equation (**20**). Similarly, when $H<\hat{h}_n$, the depth changes by Δh at a distance seaward from the coast given by

$$x=\left(\frac{h_c^3-h_n^3}{h_c^3}\right)\frac{\Delta h}{k}. \qquad \text{(21)}$$

Just upstream of the jump h equals $\hat{H}$, so the distance of the jump from the coast is obtained by putting Δh equal to $\hat{H}-h_n$ in equation (**21**). These relationships will be used to calculate jump positions in Section VI.

The various possible types of flow which can occur near the foot of a slope can be summarized as follows :

(*a*) $F_n<1$	No jump occurs
(*b*) $F_n>1,\ H>\hat{h}_n$	A jump occurs landward of the coast
(*c*) $F_n>1,\ H=\hat{h}_n$	A jump occurs at the coast
(*d*) $F_n>1,\ H<\hat{h}_n$	A jump occurs seaward of the coast.

The condition for type (*d*) is also the condition for strong winds to occur at the coast.

V. THE EFFECT OF THE EARTH'S ROTATION

The general equations governing the steady motion of a layer of cold air down a slope, taking account of the Earth's rotation, can be written

$$\frac{\partial}{\partial x}(hu)+\frac{\partial}{\partial y}(hv)=0, \qquad \text{(22)}$$

$$\frac{\partial}{\partial x}(hu^2)+\frac{\partial}{\partial y}(huv)=\frac{h\theta' g}{\theta}\left(\alpha-\frac{\partial h}{\partial x}\right)-hvl-kVu, \qquad \text{(23)}$$

$$\frac{\partial}{\partial x}(huv)+\frac{\partial}{\partial y}(hv^2)=-\frac{h\theta' g}{\theta}\frac{\partial h}{\partial y}+hul-kVv, \qquad \text{(24)}$$

where y is the cross-slope coordinate, l is the Coriolis parameter, v is the cross-slope component of wind, and V is the wind speed, u being the down-slope velocity as before. No attempt will be made to find the general solution of this system of non-linear partial differential equations. Certain particular solutions can, however, be found very easily and two of these will be discussed briefly below.

To determine a solution appropriate to the Antarctic ice slope, cross-slope variations can be neglected, hu is then equal to a constant, as before, and the equations reduce to

$$Q\frac{\mathrm{d}u}{\mathrm{d}x}=\frac{h\theta' g}{\theta}\left(\alpha-\frac{\mathrm{d}h}{\mathrm{d}x}\right)-hvl-kVu, \quad \dots\dots\dots\dots \quad (25)$$

$$Q\frac{\mathrm{d}v}{\mathrm{d}x}=Ql-kVv. \quad \dots\dots\dots\dots\dots\dots\dots\dots \quad (26)$$

Multiplying (**25**) by u and (**26**) by v and adding gives

$$Q\frac{\mathrm{d}(\frac{1}{2}V^2)}{\mathrm{d}x}=+\frac{\theta' gQ}{\theta}\left(\alpha-\frac{\mathrm{d}h}{\mathrm{d}x}\right)-kV^3. \quad \dots\dots\dots\dots \quad (27)$$

When the flow is uniform it can be seen immediately from this equation that the normal speed V_n is given by

$$V_n^3=\alpha\theta' gQ/k\theta. \quad \dots\dots\dots\dots\dots\dots \quad (28)$$

The speed of a uniform katabatic wind is therefore independent of the Earth's rotation. The air, however, no longer flows directly down the slope, consequently the down-slope component of the wind is reduced. It is this component which is relevant in determining the position of the jump when it is parallel to the coast; the position of the jump will therefore depend to a certain extent on rotational effects.

Having determined V_n it is simplest to express the other flow characteristics in terms of it. Thus

$$\sin\beta=\frac{V_n l}{g\alpha\theta'/\theta}, \quad \dots\dots\dots\dots\dots\dots \quad (29)$$

where β is the angle through which the flow is deflected by the Earth's rotation. This angle is determined by the ratio of the geostrophic acceleration $V_n l$ to the katabatic acceleration $g\alpha\theta'/\theta$, these quantities often being of comparable magnitude. The normal depth is given by

$$h_n=\frac{Q}{V_n\cos\beta}=\frac{Q}{V_n\{1-(V_n l\theta/g\alpha\theta')^2\}^{\frac{1}{2}}}. \quad \dots\dots\dots \quad (30)$$

In this type of flow the deflection from the lines of greatest slope is such that the component of the katabatic acceleration perpendicular to the lines of flow balances the geostrophic acceleration. If $V_n l>g\alpha\theta'/\theta$ then no such balance can be set up, however great the deflection of the flow. In these circumstances uniform flow is impossible. This condition can be written

$$Q>(g\alpha\theta'/\theta)^2(k/l^3)=Q_0. \quad \dots\dots\dots\dots\dots\dots \quad (31)$$

If $Q_0 \gg Q$, then rotation does not greatly affect the uniform flow. If Q_0 only just exceeds Q, then h_n is large and β is close to $\frac{1}{2}\pi$, indicating that the air tends to flow across the slope rather than down it.

The conditions under which the various types of flow occur, as listed at the end of Section IV, are altered by the introduction of rotation. As will be shown subsequently, these changes are only small in the neighbourhood of Commonwealth Bay but could be important under other conditions. The relevant Froude number for uniform flow is now $\alpha \cos^3 \beta/k$ instead of α/k, and the condition for a jump to occur parallel to the coast somewhere in the flow is $\alpha \cos^3 \beta/k > 1$. Therefore rotation always reduces the likelihood of a jump occurring. The condition for strong winds at the coast, i.e. the condition for flow type (*d*), is now

$$H < \tfrac{1}{2}h_n[(1+8\alpha \cos^3 \beta/k)^{\frac{1}{2}}-1].$$

It can be seen that rotation decreases the quantity in the square brackets and, from equation (**30**), that it increases h_n. It is not clear from these simple considerations whether rotation will increase or decrease the tendency for strong winds at the coast. A more detailed analysis indicates that the tendency for strong winds is decreased for all values of β and α/k.

The preceding solution applies only to the free descent of air down a slope. If the air is constrained to flow in a certain direction by local topography, then the solution will no longer apply. Suppose, for instance, that the air is constrained to flow directly down slope, then, under uniform flow conditions, equations (**23**) and (**24**) reduce to

$$h_n\frac{\theta'\alpha g}{\theta}-ku_n^2=0, \qquad \text{(32)}$$

$$-\frac{\theta' g}{\theta}\frac{dh}{dy}+u_n l=0. \qquad \text{(33)}$$

The solution is

$$h_n=\frac{l^2\alpha\theta}{4k\theta' g}y^2,$$

$$u_n=\frac{l}{2}\frac{\alpha}{k}y,$$

giving

$$F_n=\frac{\alpha}{k}.$$

Therefore, although h_n and u_n vary across the slope, the Froude number remains constant. The geostrophic acceleration is here balanced by the change in depth across the slope, the air tending to pile up against whatever obstacle is constraining the flow.

VI. Comparison with Observation

In order to determine whether the katabatic wind near Commonwealth Bay is shooting or tranquil we may take as typical values: $h_n=300$ m, $u_n=30$ m sec^{-1}, and $\theta'/\theta=1/60$, giving a Froude number of 18. There is consequently little doubt that the katabatic flow is shooting and a jump must therefore occur somewhere in the neighbourhood of the coast. The general explanation of the curious behaviour of the wind is now clear. The normal condition of strong steady winds experienced at the coast corresponds to flow of type (*d*). If the jump moves across the coast in response to an increase in H or a decrease in h_n, then lull conditions will prevail, corresponding to flow of type (*b*). Madigan's account of conditions during a lull, with strong winds both audible and visible higher up the slope, describes exactly what one would expect near a pressure jump (see Fig. 3).

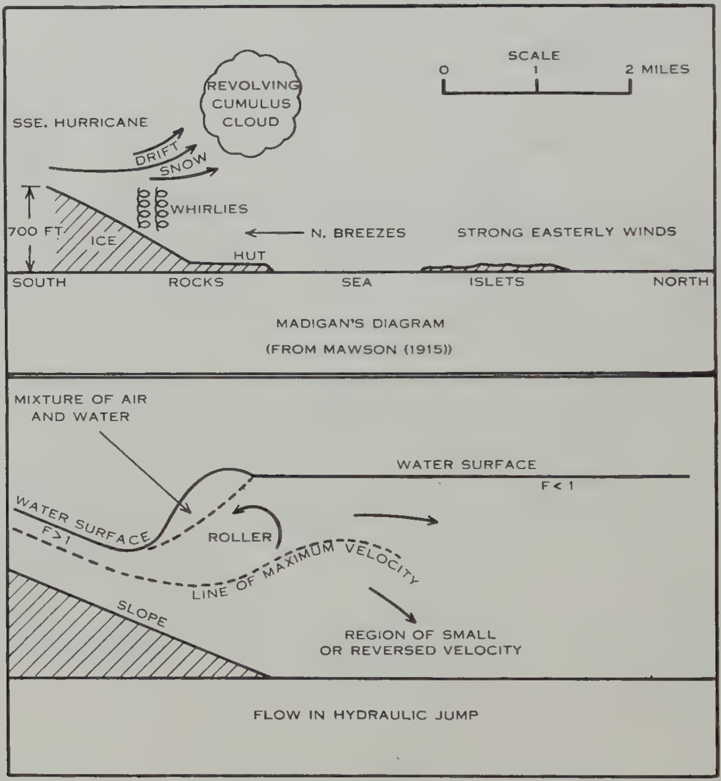


Fig. 3.—Comparison between Madigan's diagram and flow in a hydraulic jump.

As there is a considerable degree of uncertainty in the correct value for θ'/θ, we will take two values $0{\cdot}03$ and $0{\cdot}015$ corresponding to values for F_n of 10 and 20 respectively. In Commonwealth Bay α is about 10^{-1}. With these assumptions the following results are obtained :

F_n	k	h_c (m)	$\hat{h}_n$ (m)	Δp (mb)
10	10^{-2}	650	1200	$3{\cdot}2$
20	5×10^{-3}	800	1900	$2{\cdot}9$

Δp is the intensity of the pressure jump when it is situated at the coast. These values of k may be compared with a value of about 2×10^{-3} deduced from Liljequist's (1953) values for the roughness length on Antarctic ice. It is to be expected that the k value for the katabatic wind will be higher than the k value for surface friction alone since the wind is presumably retarded at its upper as well as its lower boundary. Near Commonwealth Bay the occurrence of strong winds at the coast indicates that the characteristic inversion height over the sea is generally less than $\hat{h}_n$. This is in agreement with a measured height of about 1000 m off the coast (Jalu 1950). The calculated pressure change is in good agreement with the observed change of about 3 mb. Furthermore, the pressure change at the commencement of a lull, when the jump is moving inland, is generally greater than the change at the end of a lull, when the jump is moving seawards (Boujon 1954). This confirms the prediction based on the theory of the moving jump (Section III).

If H exceeds $\hat{h}_n$ by 50 m then the jump will be situated about $\frac{1}{2}$ km inland (see equation (**20**)), whereas, if H is 50 m less than $\hat{h}_n$, the jump will be situated 2–4 km from the coast. Thus the seaward movement in response to a decrease in H is greater than the landward movement in response to a similar increase in H. This suggests that the jump will generally be most nearly stationary just inland of the coast and conditions there will be most suitable for the observational study of the jump.

The deviation of the wind from the lines of greatest slope caused by the Earth's rotation is 16° ($\theta'/\theta=0\cdot015$) and 8° ($\theta'/\theta=0\cdot03$) from equation (**29**). This compares with 10–15° observed at Cape Denison and about 5° observed at Port Martin. After passing through the jump the inclination will, according to equation (**16**), change from 16 to 60° in the first case and 8 to 30° in the second case. It is difficult to compare this with observation because the amount of turbulence generated at the jump is so great that typical turbulent velocities exceed the mean velocity downstream of the jump. It follows that the wind direction is extremely variable in that region, though further out to sea it generally settles down to an easterly. On the main Antarctic ice slope the principal effect of rotation is a slight deviation of the wind. Q_0 is about 10^5 m^2 sec^{-1} and probably exceeds any rate of flow that occurs in practice. However, on gentler slopes of equal or greater smoothness, with small inversion strengths, Q may quite often be greater than Q_0 and uniform katabatic winds will not then occur. For example, if α is 10^{-2} and the roughness and inversion strength remain the same as before, then Q_0 is about 10^3 m^2 sec^{-1}. This compares with the value for Q of about 10^4 m^2 sec^{-1} observed near Commonwealth Bay.

The predicted cross-slope velocity for uniform flow may be quite considerable, whereas the maximum cross-slope velocity that can develop when the air moves a distance x down slope is given by lx, i.e. $0\cdot15$ m sec^{-1} per km of movement. Thus in order to attain a velocity of 15 m sec^{-1} the air must move a distance of 100 km. In view of this it seems unlikely that uniform flow will be fully developed where the predicted cross-slope velocity is large. This idea can easily be translated into more precise terms. The required condition is

$$lx>v_n=V_n \sin \beta.$$

Substitution for V_n and $\sin\beta$ from equations (**28**) and (**29**) gives

$$x>(Q^2\theta/k^2\alpha g\theta')^{1/3}.$$

Whence, from equation (5),

$$x > h_c/(k^2\alpha)^{1/3} = \lambda \text{ say.}$$

An interesting feature of this inequality is the disappearance of the factor depending on the Earth's rotation. The values of λ corresponding to $\theta'/\theta = 0 \cdot 03$ and $\theta'/\theta = 0 \cdot 015$ are 30 and 60 km respectively. In the latter case the slope is probably not long enough for the development of uniform flow and the deviation near the foot of the slope will on these grounds be less than predicted on the assumption of uniform flow. In addition changes in the direction of the lines of greatest slope are liable to cause deviations which persist for considerable distances down slope. The observed deviation at any point will depend not only on the influence of the Earth's rotation but also on the irregularities of the slope upwind.

VII. References

Benjamin, T. B., and Lighthill, M. J. (1954).—*Proc. Roy. Soc.* A **224** : 448.

Binnie, A. M., and Orkney, J. C. (1955).—*Proc. Roy. Soc.* A **230** : 237.

Boujon, H. (1954).—Les Obs. Mét. de Port Martin en Terre Adélie. Fascicule I.

Cornish, V. (1934).—" Ocean Waves." p. 92. (Cambridge Univ. Press.)

Defant, A. (1933).—*Ber. Preuss. Akad. Wiss. Phys. Math.* **18** : 624.

Jalu, R. (1950).—Exp. Antarctique Française 1948–49, Premier Exposé des Résultats.

Jeffreys, H. (1925).—*Phil. Mag.* **49** : 793.

Liljequists, G. H. (1953).—Proc. Toronto Met. Conf., p. 78.

Madigan, C. T. (1929).—Sci. Rep. Aust. Antarct. Exped. 1911–14 B, Vol. 4.

Mawson, Sir Douglas (1915).—" Home of the Blizzard." Vol. 1. (Heinemann : London.)

Mirrlees, S. T. A. (1934).—Geophys. Mem. Lond. No. 61.

Prudhomme, A., and Boujon, A. H. (1952).—*Météorologie* **1952** : 132.

Woodward, F. M., and Posey, C. J. (1941).—" Hydraulics of Steady Flow in Open Channels." (Wiley & Sons : New York.)

THE CONDITION OF REGULARITY IN SIMPLE MARKOV CHAINS

By J. Gani*

[*Manuscript received December* 29, 1955]

Summary

The paper generalizes a proof, and outlines an alternative to it, for the well-known theorem on the conditions of regularity in a simple Markov chain; this is that the necessary and sufficient conditions for a chain to be regular are that the latent root 1 of the stochastic matrix for the chain must be simple, and the remaining roots have moduli less than 1.

I. Introduction

A simple Markov chain with s possible states $E_1, \ldots, E_s$, is defined by the stochastic matrix of transition probabilities

$$\mathbf{p}=\{p_{ij}\} \quad (i, j=1, \ldots, s),$$

where $p_{ij}=Pr(E_j \mid E_i) \geqslant 0$. Since, in a one-step transition from E_i, one of the states $E_i, \ldots, E_s$ must be reached, it follows that the probabilities in any row i of the stochastic matrix must sum up to 1, so that

$$\sum_{j=1}^{s} p_{ij}=1.$$

This property leads to the interesting result that 1 is a latent root of any stochastic matrix $\mathbf{p}$; for if we write

$$D(\mu)=|\mu\mathbf{I}-\mathbf{p}|, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

we find that on adding all columns of the determinant to the first,

$$D(\mu)=\begin{vmatrix} \mu-1 & -p_{12} & \ldots & -p_{1s} \\ \mu-1 & \mu-p_{22} & \ldots & -p_{2s} \\ . & . & \ldots & . \\ \mu-1 & -p_{s2} & \ldots & \mu-p_{ss} \end{vmatrix},$$

from which it is clear that $\mu_1=1$ is a latent root of the characteristic equation $D(\mu)=0$. It is, in addition, possible to show that all the latent roots μ_r $(r=1, \ldots, s)$ of $\mathbf{p}$ have moduli $|\mu_r| \leqslant 1$; for if $D(\mu)=0$, there exists at least one non-trivial vector solution $\mathbf{R}(\mu)$ of the matrix equation

$$(\mu\mathbf{I}-\mathbf{p})\mathbf{R}(\mu)=0, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

---

* Australian National University, Canberra, A.C.T.; present address: Statistical Laboratory, University of Manchester, England.

where $\mathbf{R}(\mu)$ may have real or complex elements. If $R_i(\mu)$ is the element in $\mathbf{R}(\mu)$ with the largest modulus M, then from (**2**), since $\sum_{j=1}^{s} p_{ij}R_j=\mu R_i$, it follows that

$$|\mu| M \leqslant \sum_{j=1}^{s} p_{ij} |R_j| \leqslant M,$$

so that for all values of $r=1, 2, \ldots, s$, the modulus $|\mu_r| \leqslant 1$. Both these results are given by Fréchet (1938, p. 105).

Let us now consider the probability $p_{ij}^{(2)}$ of a transition from E_i to E_j in exactly two steps; it is easy to see, by taking into account all the possible ways in which this can occur, that

$$p_{ij}^{(2)}=\sum_{r=1}^{s} p_{ir}p_{rj}.$$

Similarly, it will follow that if $p_{ij}^{(n)}$ is the higher transition probability that E_j be reached from E_i in exactly n steps, that

$$p_{ij}^{(n)}=\sum_{r=1}^{s} p_{ir}p_{rj}^{(n-1)} \quad (n=2, \ldots).$$

Clearly, since one of the states $E_1, \ldots, E_s$ must necessarily be reached starting from E_i, then

$$\sum_{j=1}^{s} p_{ij}^{(n)}=1,$$

and the $p_{ij}^{(n)}$ are (i, j)th elements of a stochastic matrix, itself the nth power of the stochastic matrix $\mathbf{p}$, so that

$$\mathbf{p}^n=\{p_{ij}^{(n)}\}.$$

The Markov chain known as regular is that for which

$$\lim_{n\to\infty} p_{ij}^{(n)}=P_j \geqslant 0 \quad (j=1, \ldots, s), \quad \ldots\ldots\ldots\ldots\ (\mathbf{3})$$

where P_j is some fixed finite stationary probability independent of the initial state E_i. This result can be written in matrix form as

$$\lim_{n\to\infty} \mathbf{p}^n=\mathbf{1P}', \quad \ldots\ldots\ldots\ldots\ldots\ (\mathbf{4})$$

where $\mathbf{P}'$ is the row vector of probabilities P_j, and $\mathbf{1}$ is the column vector of unit elements. It is obvious that $\sum_{j=1}^{s} P_j=1$, so that at least one of these stationary probabilities is non-zero. Since

$$\mathbf{1P}'=\lim_{n\to\infty} \mathbf{p}^{n+1}=\lim_{n\to\infty} \mathbf{p}^n\mathbf{p}=\mathbf{1P}'\mathbf{p},$$

the row vector $\mathbf{P}'$ of probabilities P_j is a solution of the matrix equation

$$\mathbf{P}'=\mathbf{P}'\mathbf{p}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ (\mathbf{5})$$

and such a vector can easily be shown to be unique.

It is the purpose of the present paper to extend a known but not quite general proof, and to give an elementary alternative to this generalization, for the well-known theorem on the regularity of a Markov chain:

> The necessary and sufficient conditions for a simple Markov chain to be regular, that is, for $\lim_{n\to\infty} p_{ij}^{(n)} = P_j$, are that the latent root 1 of the stochastic matrix for the chain be simple, and the remaining roots have moduli less than 1.

It is comparatively simple to prove this by considering higher transition probabilities in the case where the latent roots of **p** are assumed to be all simple, whether as in Bartlett (1955, Section 2.2) by the method of spectral resolution, or as in Feller (1950, Section 16.1) by using a partial fractions expansion. Fréchet has given a proof of the theorem in the general case when some of the latent roots may be multiple (1938, pp. 109–14), but this is not direct, arising as it does in the study of higher transition probabilities for various types of Markov chains. We briefly outline Bartlett's proof, give a generalization of it in the case where the latent roots of **p** may be multiple, which depends on a little-used theorem in canonical matrices, and finally provide an elementary alternative to this.

II. The Generalization of Bartlett's Proof

In essence, Bartlett's (1955) proof of the necessary conditions for the regularity of a Markov chain depends on the fact that, if the stochastic matrix **p** of the chain has latent roots, all of which are simple, it can be expressed as

$$\mathbf{p} = \mathbf{T}^{-1}\mathbf{u}\mathbf{T},$$

where **T** is a non-singular matrix, and $\mathbf{u} = \{\mu_j \delta_{ij}\}$ is the diagonal matrix whose elements are the latent roots μ_j $(j=1, \ldots, s)$ of **p**. It is clear then, if equation (3) or its matrix equivalent (4) are to hold, that

$$\mathbf{p}^n = \mathbf{T}^{-1}\mathbf{u}^n\mathbf{T} \to \mathbf{1P}',$$

so that

$$\lim_{n\to\infty} \mathbf{u}^n = \mathbf{T1P}'\mathbf{T}^{-1}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

where the elements of $\mathbf{T1P}'\mathbf{T}^{-1}$ have fixed finite values. If therefore the unit root is $\mu_1 = 1$ say, it is clear that the remaining roots μ_r $(r=2, \ldots, s)$ are necessarily such that $|\mu_r| < 1$.

The generalization of this in the case where the roots of **p** may be multiple depends on the theorem in canonical matrices (Turnbull and Aitken 1932, Section 6) which states that if $\mu_1, \ldots, \mu_t$ are latent roots of **p** of multiplicity $m_1, \ldots, m_t$ respectively, where $\sum_{r=1}^{t} m_r = s$, then the stochastic matrix can be expressed as

$$\mathbf{p} = \mathbf{T}^{-1}\mathbf{MT} = \mathbf{T}^{-1}\begin{bmatrix} \mathbf{M}_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & \mathbf{M}_t \end{bmatrix}\mathbf{T},$$

where $\mathbf{T}$ is again non-singular, and the canonical matrix $\mathbf{M}$ is partitioned into t diagonal matrices $\mathbf{M}_1, \ldots, \mathbf{M}_t$ of orders $m_1, \ldots, m_t$ respectively, each matrix $\mathbf{M}_i$ consisting of one or more non-overlapping diagonal submatrices $\mathbf{C}_i^{(v)}$ of the "simple classical" type

$$\mathbf{C}_i^{(v)} = \begin{bmatrix} \mu_i & 1 & 0 & \cdots & \cdot & 0 \\ 0 & \mu_i & 1 & 0 \cdots & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdots 0 & \mu_i & 1 \\ 0 & \cdot & \cdot & \cdots & 0 & \mu_i \end{bmatrix},$$

where $v \leqslant m_i$ indicates the order. For example, if the root μ_i has multiplicity $m_i = 6$, a possible form for the matrix $\mathbf{M}_i$ could be

$$\mathbf{M}_i = \begin{bmatrix} [\mu_i] & & \\ & \begin{bmatrix} \mu_i & 1 & 0 \\ 0 & \mu_i & 1 \\ 0 & 0 & \mu_i \end{bmatrix} & \\ & & \begin{bmatrix} \mu_i & 1 \\ 0 & \mu_i \end{bmatrix} \end{bmatrix}.$$

If $m_i = 1$, the matrix $\mathbf{M}_i$ must consist of the single diagonal element μ_i.

In this case again

$$\mathbf{p}^n = \mathbf{T}^{-1}\mathbf{M}^n\mathbf{T} \to \mathbf{1P}',$$

so that

$$\lim_{n\to\infty} \mathbf{M}^n = \mathbf{T1P}'\mathbf{T}^{-1}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

where, just as before, the elements of $\mathbf{T1P}'\mathbf{T}^{-1}$ are fixed finite values. Now suppose we raise a simple classical matrix $\mathbf{C}_i^{(v)}$ to the nth power; we obtain for $n \geqslant v-1$

$$\{\mathbf{C}_i^{(v)}\}^n = \begin{bmatrix} \mu_i^n & n\mu_i^{n-1} & \cdot & \cdots & \cdot & \binom{n}{v-1}\mu_i^{n-v+1} \\ \cdot & \mu_i^n & n\mu_i^{n-1} & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdots 0 & \mu_i^n & n\mu_i^{n-1} \\ 0 & \cdot & \cdot & \cdots & 0 & \mu_i^n \end{bmatrix},$$

where the elements in the main diagonal, and the diagonal above it are respectively

$$\mu_i^n, \binom{n}{1}\mu_i^{n-1}, \binom{n}{2}\mu_i^{n-2}, \ldots, \binom{n}{v-1}\mu_i^{n-v+1},$$

and the remaining elements are zero.

If the unit root is $\mu_1 = 1$ say, then it is clear again that if (7) is to hold, $|\mu_r| < 1$ $(r = 2, \ldots, t)$; further, even if the multiplicity of μ_1 is $m_1 > 1$, it is not possible that $\mathbf{M}_1$ consist of simple classical matrices of order greater than 1, so that at the most the form of $\mathbf{M}_1$ is

$$\mathbf{M}_1 = \begin{bmatrix} [1] & & & & \\ & [1] & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & [1] \end{bmatrix} = \mathbf{I},$$

of order $m_1 > 1$. In this case $\mathbf{M}_1^n = \mathbf{I}$, and we have from (7) that

$$\lim_{n \to \infty} \mathbf{M}^n = \begin{bmatrix} \mathbf{I} & & & & \\ & \mathbf{O} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & \mathbf{O} \end{bmatrix} = \mathbf{T} \begin{bmatrix} P_1 & \cdot & \cdot & \cdot & P_s \\ P_1 & \cdot & \cdot & \cdot & P_s \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ P_1 & \cdot & \cdot & \cdot & P_s \end{bmatrix} \mathbf{T}^{-1}.$$

Now the rank of $\mathbf{I}$ is m_1; but since $\mathbf{T}$ is a non-singular matrix, the rank of the right-hand matrix will be the same as that of $\mathbf{1P}'$, that is 1. It follows that $m_1 = 1$, so that the root $\mu_1 = 1$ is simple.

III. An Elementary Proof of the Theorem

The following proof, which is equivalent to the preceding one, is based on a simple way of obtaining the expression for the higher transition probability $p_{ij}^{(n)}$. To do this, we shall consider for some suitable value of μ, the expansion in powers of $\mathbf{p}$ of the matrix

$$(\mu\mathbf{I} - \mathbf{p})^{-1} = \mu^{-1}(\mathbf{I} + \mu^{-1}\mathbf{p} + \mu^{-2}\mathbf{p}^2 + \ldots);$$

this is convergent for $|\mu| > 1$. It is clear that the (i, j)th element of this matrix is

$$\{(\mu\mathbf{I} - \mathbf{p})^{-1}\}_{ij} = \mu^{-1}(\delta_{ij} + \mu^{-1}p_{ij} + \mu^{-2}p_{ij}^{(2)} + \ldots),$$

so that, if $\{(\mu\mathbf{I} - \mathbf{p})^{-1}\}_{ij}$ can be expanded in some other way in descending powers of μ, it is possible to obtain an expression for $p_{ij}^{(n)}$ by equating the coefficients of $\mu^{-(n+1)}$ in the two expansions.

Consider the matrix

$$(\mu\mathbf{I} - \mathbf{p}') = \begin{bmatrix} \mu - p_{11} & -p_{21} & \cdot\cdot\cdot & -p_{s1} \\ -p_{12} & \mu - p_{22} & \cdot\cdot\cdot & -p_{s2} \\ \cdot & \cdot & \cdot\cdot\cdot & \cdot \\ -p_{1s} & \cdot & \cdot\cdot\cdot & \mu - p_{ss} \end{bmatrix},$$

the cofactors $C_{ij}(\mu)$ of each of the elements $(\mu\delta_{ji} - p_{ji})$ of the matrix are of degree no greater than $s-1$ in μ, whereas the degree of $D(\mu) = |\mu\mathbf{I} - \mathbf{p}|$ is always one greater than that of any C_{ij}, that is, no greater than s. It follows that

$$\{(\mu\mathbf{I} - \mathbf{p})^{-1}\}_{ij} = C_{ij}(\mu)\,|\mu\mathbf{I} - \mathbf{p}|^{-1} = C_{ij}(\mu)/\{(\mu - \mu_1)^{m_1} \ldots (\mu - \mu_t)^{m_t}\}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

where $\mu_1, \ldots, \mu_t$ are the roots of the characteristic equation $D(\mu)=0$, with multiplicity $m_1, \ldots, m_t$ respectively, where $\sum_{r=1}^{t} m_r = s$.

From (8), by the method of partial fractions, we obtain that

$$\{(\mu\mathbf{I}-\mathbf{p})^{-1}\}_{ij}=\sum_{r=1}^{t} B_r^{(ij)}(\mu)/(\mu-\mu_r)^{m_r}, \quad \ldots\ldots\ldots\ldots \quad (9)$$

where the $B_r^{(ij)}(\mu)$ are polynomials in μ of degree no greater than m_r-1, of the form

$$B_r^{(ij)}(\mu)=\sum_{u=0}^{m_r-1} b_{ru}^{(ij)}\mu^u\,;$$

obviously, if $m_r=1$, $B_r^{(ij)}(\mu)=b_{r0}^{(ij)}$. Since $|\mu|>1$, and we know that for any stochastic matrix $\mathbf{p}$, $|\mu_r|\leqslant 1$ for all r, we may expand the expression (9) as follows:

$$\{(\mu\mathbf{I}-\mathbf{p})^{-1}\}_{ij}=$$
$$\sum_{r=1} \mu^{-m_r}\left\{\sum_{u=0}^{m_r-1} b_{ru}^{(ij)}\mu^u\right\}\left\{1+m_r\mu^{-1}\mu_r+\ldots+\binom{m_r+n-1}{n}\mu^{-n}\mu_r^n+\ldots\right\}. \quad \ldots\ldots\ldots\ldots \quad (10)$$

Let m be the largest of $m_1, \ldots, m_t$; then, for all values of $n\geqslant m-1$, the coefficient of $\mu^{-(n+1)}$ in the element $\{(\mu\mathbf{I}-\mathbf{p})^{-1}\}_{ij}$ of (10) is equal to

$$\sum_{r=1}^{t}\left\{b_{r,m_r-1}^{(ij)}\mu_r^n\binom{m_r+n-1}{n}+\ldots+b_{r0}^{(ij)}\mu_r^{n-m_r+1}\binom{n}{n-m_r+1}\right\}, \quad \ldots \quad (11)$$

or for simplicity,

$$\sum_{r=1}^{t} \beta_r^{(ij)}(n)\mu_r^n, \quad \ldots\ldots\ldots\ldots \quad (12)$$

where $\beta_r^{(ij)}(n)$ are polynomials in n of degree no greater than m_r-1.

It is clear that this is the transition probability $p_{ij}^{(n)}$ for values of n greater than or equal to $m-1$; this is all that is required, since we are to consider the limit of $p_{ij}^{(n)}$ as $n\rightarrow\infty$. Writing, for sufficiently large n, the result

$$p_{ij}^{(n)}=\sum_{r=1}^{t} \beta_r^{(ij)}(n)\mu_r^n, \quad \ldots\ldots\ldots\ldots \quad (13)$$

we see that, if the chain is regular, so that the condition (3) holds, then if the unit root is $\mu_1=1$ say, all the remaining roots whether simple or multiple must have moduli $|\mu_r|<1$ $(r=2, \ldots, t)$; further μ_1 cannot be multiple, for if it were, then

$$\lim_{n\rightarrow\infty} p_{ij}^{(n)}=\lim_{n\rightarrow\infty} \beta_1^{(ij)}(n)=\infty,$$

which contradicts (3). It follows that the root $\mu_1=1$ is simple, the remaining roots having moduli $|\mu_r|<1$ $(r=2, \ldots, t)$.

The sufficient condition for regularity presents no difficulty; it is that, if the unit root of the stochastic matrix $\mathbf{p}$ is simple, and the remaining roots have moduli less than 1, then $\lim_{n\to\infty} p_{ij}^{(n)}=P_j$ as in (**3**). From the result (**13**), it is clear that if $\mu_1=1$ is simple, and $|\mu_r|<1$ $(r=2, \ldots, t)$, then

$$\lim_{n\to\infty} p_{ij}^{(n)}=b_{10}^{(ij)},$$

where $b_{10}^{(ij)}$ is some constant. In matrix form this can be written

$$\lim_{n\to\infty} \mathbf{p}^n=\mathbf{b}=\begin{bmatrix} \mathbf{b}^{(1)} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{b}^{(s)} \end{bmatrix}, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{14})$$

where $\mathbf{b}$ is the matrix of elements $b_{10}^{(ij)}$, and $\mathbf{b}^{(i)}$ its row vectors with elements such that $\sum_{j=1}^{s} b_{10}^{(ij)}=1$. We show that these are identically equal to a single row vector $\mathbf{P}'$ with elements P_j such that $\sum_{j=1}^{s} P_j=1$. For, from (**14**) we have that

$$\mathbf{b}=\lim_{n\to\infty} \mathbf{p}^{n+1}=\lim_{n\to\infty} \mathbf{p}^n\mathbf{p}=\mathbf{bp},$$

so that for each $\mathbf{b}^{(i)}$ $(i=1, \ldots, s)$,

$$\mathbf{b}^{(i)}\mathbf{p}=\mathbf{b}^{(i)}.$$

But, since the latent root $\mu_1=1$ of $\mathbf{p}$ is simple, then there exists only one row vector $\mathbf{P}'$ of probabilities P_j such that $\sum_{j=1}^{s} P_j=1$, which is a solution of the equation (**5**) (Frazer, Duncan, and Collar 1947, Section 3.5). It follows that

$$\mathbf{b}^{(i)}=\mathbf{P}' \quad (i=1, \ldots, s),$$

so that $\lim_{n\to\infty} p_{ij}^{(n)}=P_j$ as required.

IV. ACKNOWLEDGMENT

The author is indebted to Dr. E. J. Hannan for his helpful criticism and suggestions.

V. REFERENCES

Bartlett, M. S. (1955).—" An Introduction to Stochastic Processes." (Cambridge Univ. Press.)

Feller, W. (1950).—" An Introduction to Probability Theory and Its Applications." (John Wiley: New York.)

Frazer, R. A., Duncan, W. J., and Collar, A. R. (1947).—" Elementary Matrices." (Cambridge Univ. Press.)

Fréchet, M. (1938).—" Recherches Théoriques Modernes." Vol. 2. (Gauthier-Villars: Paris.)

Turnbull, H. W., and Aitken, A. C. (1932).—" An Introduction to the Theory of Canonical Matrices." (Blackie: London.)

STATISTICAL ANALYSIS OF AUSTRALIAN PRESSURE DATA

By S. C. Das*

[*Manuscript received February* 8, 1956]

Summary

We give a statistical analysis of the pressure data along the east coast of Australia, to test a hypothesis which arose out of Deacon's (1953) work, that there was a shift in the mean high pressure belt. It is shown that the belt is slowly moving southward.

I. Introduction

In the present paper we consider the statistical analysis of Australian pressure data in order to test the following hypothesis which arose out of Deacon's (1953) work. Deacon compared the mean daily maximum temperatures for the summer seasons between two 30 year periods 1881–1910 and 1911–40 for some Australian inland localities. Table 1 of his paper gives temperature differences for the two periods for 14 stations. All these show a consistently lower temperature in period 2 than in period 1. He also found that for the period 1911–50, the summer rainfall over much of the southern part of Australia was considerably greater than in the previous 30 years. He conjectured that these changes were due to a climatic trend, and by noting the difference in the character of the annual variation in atmospheric pressure between these two periods he suggested that the cause of the phenomena might be a shift in the mean high pressure belt.

To investigate how far the conjecture is correct we shall make an analysis of the pressure data. The way in which this is done is briefly indicated below. By taking the sea-level pressure along a line across the maximum of high pressure, it is possible to determine by appropriate curve fitting the position and the intensity of the maximum at any time. Both the position and the intensity are expected to show a variation with season, but the interesting point is to examine whether any of these elements are showing a climatic trend. With this in view we proceed to determine

(1) the annual mean position of the maximum,
(2) the annual mean intensity of the maximum,
(3) the amplitude of the annual cycle of position,
(4) the amplitude of the annual cycle of intensity,

and to examine what trends, if any, are present in any one of these.

From practical considerations it was suggested that the line should incorporate the following stations: Cairns, Townsville, Rockhampton, Brisbane,

---

* Australian National University, Canberra, A.C.T.; present address: Ravenshaw College, Orissa, India.

Sydney, Moruya Heads, Gabo I., Launceston, and Hobart. Other stations for which data are available could not be included as they are either far east or west of this line and inclusion of data from them would have confused the results because of lateral pressure differences, or because they are far above sea-level and another type of error would be likely to be introduced.

II. Source of the Data

All these stations have records dating back to 1909 which provided 46 years of data. For some stations and for some years these were available from published records " Results of Rainfall Observations " made in different states. The figures given in these records are, generally, the mean of 9 a.m. and 3 p.m. pressures in inches corrected and reduced to 32 °F., M.S.L., and standard gravity. The rest of the data were supplied by the Division of Meteorological Physics, C.S.I.R.O., with a few adjustments here and there to make the values as complete and homogeneous as possible. As an example we give the adjustments made for Cairns. For a few months here, no values are available and the values for Cairns aerodrome have been inserted instead. For some other months only 9 a.m. values are available ; these were adjusted by assuming that the pressure variation between 9 a.m. and 3 p.m. at Cairns aerodrome is the same as at the post office, i.e. from the 9 a.m. values at the post office have been subtracted half the difference between the values at 9 a.m. and 3 p.m. at the aerodrome.

III. Analysis of the Data

For every month, we have the nine station values of the position (latitude) and intensity of the monthly average pressures. The first step in the analysis is to take the latitude as the independent variable and the intensity as the dependent variable, and to fit an appropriate curve to determine the intensity and position of the monthly mean maximum pressures. But the appropriate curve to be fitted depends upon the relationship between position and intensity of the monthly average pressures, which is unfortunately not known. Consequently, as a sort of pilot method, the data for a sample of months were plotted, and smooth graphs were drawn by eye through these plotted points to give some indication of this relationship. The graphs varied so much from month to month and from year to year that it was very difficult to find one or two suitable equations which would fit the data satisfactorily. The only objective method was to fit orthogonal polynomials in successive stages, the goodness of fit of consecutive terms of higher degree being observed and tested at each stage. This would have required a huge amount of labour and would not have given much more information than the subjective method followed. Besides there is the following difficulty. In certain summer months the maximum lies south of Hobart, and any reasonable curve fitted to the nine observations would not have a relative maximum, so that some subjective judgment must necessarily be used.

In view of this, the pressure data for individual months of each year for the nine given stations were plotted as ordinates against the latitudes of these stations as abscissae. By drawing smooth graphs by eye through these plotted points, the position and the intensity of the mean monthly maximum pressures were

TABLE 1

CALCULATED ANNUAL VALUES FROM METEOROLOGICAL DATA

Year	Annual Mean Position of the Maximum (deg. of lat.)	Amplitude of the Annual Cycle of Position (deg.)	Annual Mean Intensity of the Maximum (in.)	Amplitude of the Annual Cycle of Intensity (in.)
1909	31·79	2·92	30·005	0·076
1910	31·71	4·93	30·040	0·100
1911	31·50	4·49	30·028	0·107
1912	31·12	5·15	30·051	0·108
1913	31·71	5·66	30·052	0·130
1914	32·17	0·80	30·106	0·132
1915	29·21	3·40	30·010	0·084
1916	30·96	4·64	30·094	0·296
1917	30·83	5·82	30·001	0·080
1918	31·79	5·68	30·053	0·089
1919	32·12	5·54	30·078	0·111
1920	32·12	4·06	30·032	0·073
1921	32·42	4·37	30·066	0·070
1922	31·83	3·09	29·985	0·125
1923	29·12	5·74	30·010	0·094
1924	30·96	2·33	30·021	0·159
1925	32·25	4·50	30·050	0·123
1926	29·54	3·82	30·022	0·086
1927	32·00	3·88	30·123	0·230
1928	30·83	5·43	30·025	0·093
1929	30·37	4·55	30·005	0·133
1930	31·96	4·63	30·093	0·107
1931	31·25	4·15	30·050	0·097
1932	32·30	2·17	30·029	0·114
1933	34·83	2·89	30·027	0·124
1934	35·28	6·61	30·059	0·126
1935	32·50	5·08	30·014	0·100
1936	32·80	5·65	30·038	0·117
1937	34·46	2·77	30·055	0·154
1938	33·96	3·79	30·029	0·151
1939	31·58	5·63	30·033	0·082
1940	32·04	0·68	30·067	0·166
1941	31·87	3·32	30·066	0·105
1942	33·25	4·08	30·034	0·078
1943	33·04	2·02	30·005	0·090
1944	31·58	1·15	30·045	0·172
1945	32·71	4·48	30·048	0·100
1946	32·12	5·66	30·003	0·092
1947	32·83	5·45	30·048	0·105
1948	32·08	5·21	30·032	0·092
1949	32·71	4·80	30·049	0·157
1950	35·83	5·85	30·065	0·135
1951	31·75	4·81	30·028	0·096
1952	31·58	1·45	30·003	0·097
1953	31·75	3·89	30·047	0·060
1954	33·46	1·67	30·087	0·114

determined. Thus 12 values for the intensity and 12 values for the position were obtained for each of the years from 1909 to 1954.

The means (from the 46 yearly observations) for each month (for intensity and position) were then calculated. These showed a seasonal movement which was well fitted by a simple sine curve. By fitting such a sine curve the annual values for (1), (2), (3), and (4) were obtained ; these are given in Table 1.

In order to examine whether there is any trend in the annual values, a polynomial regression line of the form $Y=A+B'\xi_1'+C'\xi_2'$ has been fitted to each column of Table 1. Following the accepted notation used in Fisher and Yates's tables, ξ_1' and ξ_2' stand for orthogonal polynomials of degree one and two in x where x is the number of years measured from the centre point of the series. The results of computations are given in Table 2.

TABLE 2

REGRESSION COEFFICIENTS

Annual Values	B'	C'	$t_{B'}$	$t_{C'}$
Annual mean position of the maximum pressure ..	0·02054	—0·00044	2·94	0·19
Amplitude of the annual cycle of position ..	—0·00750	—0·00114	0·89	0·40
Annual mean intensity of the maximum	—0·00001	0·00001	0·02	0·04
Amplitude of the annual cycle of intensity ..	—0·00009	—0·00008	0·38	1·02

The first two columns give the values of the first and the second regression coefficients for each of the annual values. To test whether a regression coefficient is significantly different from zero we have computed the corresponding t by dividing the particular regression coefficient in question by its estimated standard error. These t values are given in the third and the fourth columns of the table. For 43 degrees of freedom none of the regression coefficients are significant except B' for the annual mean position of the maximum pressure. The regression equation obtained for the mean position in degrees is

$$Y=32\cdot 106+0\cdot 04108x,$$

and shows that the high pressure belt is shifting slowly towards the south. The regression coefficients for the annual mean intensity of the maximum and the amplitude of its oscillation are of course so far from significant that nothing can be said about them. While the coefficient of the first term in the regression of the amplitude of the cycle (in the position) on time is not significant, it is possible that the amplitude of the cycle is decreasing so that not only is the mean position of the maximum further south, but the oscillation about this mean position is shallower and the maximum will not tend to come as far north. Figure 1 shows the annual mean position and the above regression line.

The graph for the mean position suggests that there is, apart from the long-period trend, a certain persistence in successive values. Accordingly, we have computed the first five serial correlations of the residuals, Z_t, from the regression on the first two orthogonal polynomials. The formula used was

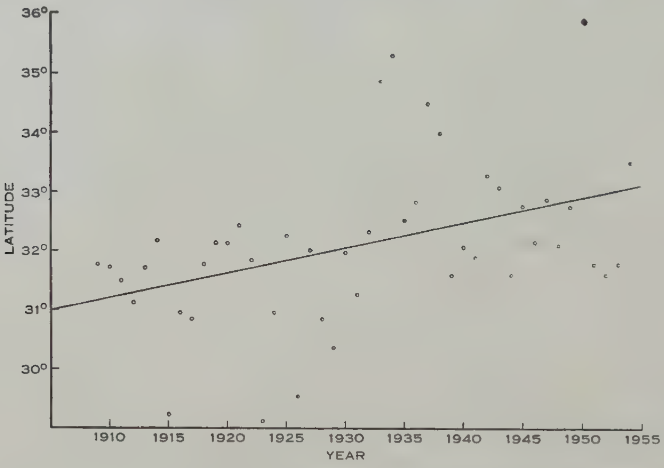


Fig. 1.—The annual mean position of the maximum pressure.

$$r_j = \sum_{1}^{46-j} Z_t Z_{t+j} \Big/ \sum_{1}^{46} Z_t^2.$$

This formula will bias the estimates downwards slightly but this will not matter for our purpose. The first five r_j's were

$$r_1 = 0\cdot226,\ r_2 = -0\cdot015,\ r_3 = 0\cdot081,\ r_4 = 0\cdot030,\ r_5 = -0\cdot010\ ;$$

clearly the last four are not significant. The first may be tested by computing

$$d = 2(1-r_1) - (Z_1^2 + Z_{46}^2) \Big/ \sum_{1}^{46} Z_t^2 = 1\cdot54,$$

and referring this to the tables of Durbin and Watson (1951). These tables give bounds for the 5 per cent. significance point for d. For a one-sided test against positive serial correlation the bounds show that the significance point lies between 1·43 and 1·62. The statistic d falls between these limits. However, Hannan (1957) has shown that for orthogonal polynomials the true significance point for d is very near to the upper bound as tabulated by Durbin and Watson. The value d is therefore significant and indicates positive serial correlation, the first serial correlation being about 0·23.

From the remaining r_j's it appears that a reasonable model of the process generating the residuals is a simple Markoff process. A test of goodness of fit of this model may be obtained by forming

$$R_j = r_j - 2r_1 r_{j-1} + r_1^2 r_{j-2}, \quad j \geqslant 2.$$

On the hypothesis of a simple Markoff process the R_j's are asymptotically normal and independent with zero mean and variance which may be estimated by

$$S^2=(1-r_1^2)^2/46.$$

Then $\chi^2_{p-1}=\sum_2^p R_j^2\Big/S^2$ is asymptotically distributed as χ^2 with $p-1$ degrees of freedom. Using the first 5 r_j's only, we obtain

$$\chi_4^2=0\cdot 748,$$

which is not significant at any reasonable level.

This test is due to Quenouille (1947).

The presence of serial correlation naturally affects the test of significance of the regression coefficients, but it is apparent in this case that the effect will not appreciably alter the significance of the first term. In fact the variance of the regression coefficients will be increased in the ratio $(1+\rho)/(1-\rho)$ (Bartlett 1935), if the true residuals are Markovian with serial correlation ρ. Even allowing for a substantial downwards bias in r_1 as an estimate of ρ, this factor is still less than 2. No exact test of significance for the regression coefficients is available when serial correlation is present in the residuals, but an approximate procedure is to treat the test statistic $t_{B'}$ as having the same type of distribution as in the case of independent residuals but with an increased variance. Using the (conservative) figure 2 for the ratio in which the variance is increased, this procedure results in the treatment of the statistic $t_{B'}$ as having come from only half the number of observations, that is 23. This would leave the first coefficient significant at the 1 per cent. point. Furthermore, Grenander (1954) has shown that, asymptotically, for the orthogonal polynomials the straightforward least square procedure is as efficient as the best linear unbiased procedure (which could be used only if the nature of the process generating the true residuals were known). Thus there is no point in recomputing the regression to take account of the serial correlation.

IV. ACKNOWLEDGMENTS

The author wishes to thank Mr. E. L. Deacon, Division of Meteorological Physics, C.S.I.R.O., for suggesting the problem; Mr. E. J. Hannan for some helpful discussions; and Mr. H. E. Gunther of the Geography Department of the Australian National University for drawing the graph.

V. REFERENCES

BARTLETT, M. S. (1935).—*J.R. Statist. Soc.* **98**: 536.
DEACON, E. L. (1953).—*Aust. J. Phys.* **6**: 209–18.
DURBIN, J., and WATSON, G. S. (1951).—*Biometrika* **38**: 159–78.
GRENANDER, U. (1954).—*Ann. Math. Statist.* **25**: 252–72.
HANNAN, E. J. (1957).—*Biometrika* (in press).
QUENOUILLE, M. H. (1947).—*J.R. Statist. Soc.* **110**: 123.

SHORT COMMUNICATIONS

SCATTERING OF HIGH ENERGY ELECTRONS BY CARBON AND THE NUCLEAR DENSITY DISTRIBUTION*

By L. J. Tassie†

The elastic scattering of 187 MeV electrons by ^{12}C has been calculated, using the Born approximation and the independent particle model with spherical box wave functions and with harmonic oscillator wave functions (Tassie 1956).

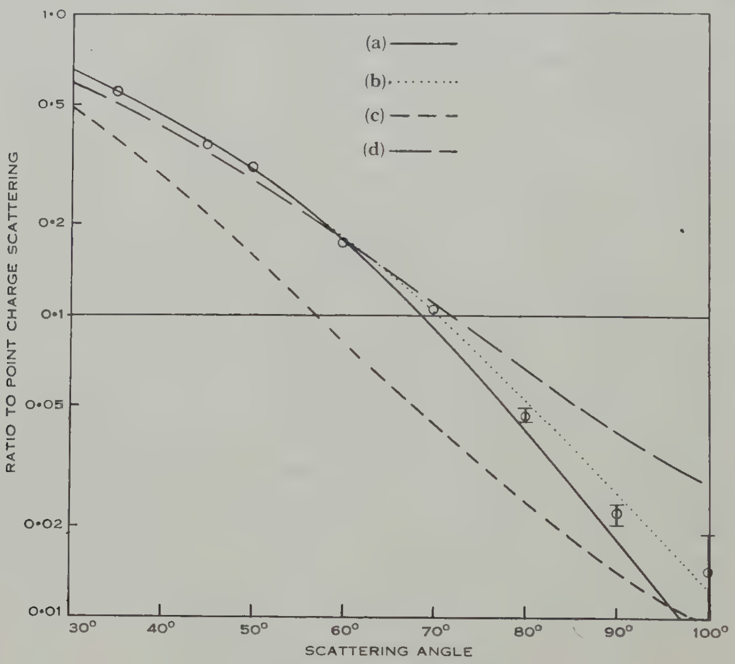


Fig. 1.—Elastic scattering of 187 MeV electrons by ^{12}C. Experimental points are those of Fregeau and Hofstadter (1955).

(*a*) Calculated using the independent particle model with spherical box wave functions.

(*b*) Calculated using the independent particle model with harmonic oscillator wave functions.

(*c*) Calculated using the nuclear density distribution of Gatha, Shah, and Patel (1954).

(*d*) Calculated using the nuclear density distribution of Gatha, Shah, and Patel reduced in linear dimensions by 0·82.

The results are plotted in Figure 1 as the ratio to point charge scattering, and are compared with the experimental results of Fregeau and Hofstadter (1955). The parameters required for best agreement with experiment were

---

* Manuscript received May 21, 1956.

† Physics Department, University of Melbourne.

$a_{SB}=3\cdot9\times10^{-13}$ cm and $a_{HO}=(h^2/4\pi^2m_pw)=1\cdot58\times10^{-13}$ cm, where w is the spacing of the harmonic oscillator energy levels.

As shown in Figure 2, the corresponding nuclear charge distributions are similar, but differ from the nuclear density distribution calculated by Gatha, Shah, and Patel (1954) from the experimental results on the scattering of 340 MeV protons by C and Al (Richardson *et al.* 1952), and their distribution when used as the charge distribution does not give an electron scattering distribution agreeing with experiment (Fig. 1) for C (Fregeau and Hofstadter 1955). While it does appear to agree with experiment for Be (Hofstadter, Fechter, and McIntyre 1953; Gatha, Patel, and Patel 1954), in this experiment, the elastic scattering was not resolved from the 2·54 MeV loss scattering, and this fact throws doubt on the agreement.

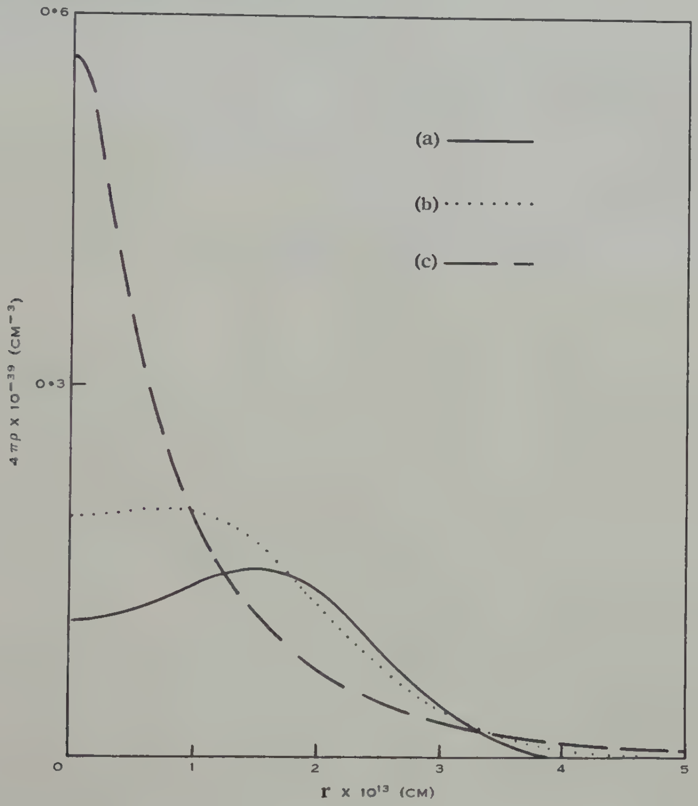


Fig. 2.—Nuclear charge and density distributions for ^{12}C. ρ normalized so that

$$\int\rho\,\mathrm{d}V=1.$$

(*a*) Using spherical box wave functions.
(*b*) Using harmonic oscillator wave functions.
(*c*) Nuclear density distribution of Gatha, Shah, and Patel (1954).

It is therefore important to examine the assumptions in the calculations of Gatha, Shah, and Patel (1954). They have used a complex potential with the

Born approximation, and this approximation is unsatisfactory for scattering by the imaginary part of the potential (Mohr and Robson 1956), although, when the imaginary potential is small, it provides a reasonable approximation for the dependence of the relative angular distribution of scattering on the shape of the scattering potential. Thus, this method should give the correct shape of the scattering potential.

It has been assumed by Gatha, Shah, and Patel (1954) that the nuclear scattering potential has the same shape as the nuclear density distribution. This assumption is valid only if the effect of the finite range of internucleon forces can be neglected. Williams (1955) has treated the problem of obtaining the nuclear scattering potential from the nucleon distribution with internucleon forces of non-zero range. Using a similar procedure, and taking the internucleon force as a square well of radius α, the nuclear potential will arise from a nucleon distribution $\rho_\alpha(r)$ given by, for small α,

$$\rho_\alpha(r)=\rho_0(r)-(\alpha^2/5r)\rho_0'(r)-(\alpha^2/10)\rho_0''(r), \quad \text{for } r\geqslant\alpha,$$

and

$$\rho_\alpha(0)=\rho_0(0)-(3\alpha^2/10)\rho_0''(0), \qquad \ldots\ldots\ldots\ldots \quad (1)$$

assuming $\rho_0'(0)=0$, where $\rho_0(r)$ is the nucleon distribution obtained by assuming internucleon forces of zero range. $\rho_\alpha(r)$ must then be normalized so that

$$\int\rho_\alpha(r)\mathrm{d}V=\int\rho_0(r)\mathrm{d}V.$$

Using the nuclear density distribution of Gatha, Shah, and Patel for $\rho_0(r)$ it was found for small α that ρ_α is more peaked at $r=0$, and of smaller radial extension than ρ_0. For ^{12}C,

$$\rho_0(r=\alpha)=0\cdot34\rho_0(0),$$

assuming a range of internucleon forces of $\alpha=10^{13}$ cm, and we can hardly expect equation (1) to be valid when the nuclear density varies so rapidly in a distance α. Nevertheless we can still expect the above conclusion for small α to apply, and the effect of the range of internucleon forces certainly cannot be neglected.

An attempt was made to fit the experimental electron scattering for ^{12}C by varying the size of Gatha, Shah, and Patel's (1954) distribution, but it is impossible to fit the scattering at all angles because the distribution is too peaked at the origin. Figure 1 shows the scattering for a reduction in size of this distribution of 0·82.

Since by taking the finite range of nuclear forces into account a more peaked nucleon distribution would be derived from the nucleon scattering experiments, there would seem to be some discrepancy in the shape of the nucleon distribution for light nuclei indicated by high energy electron scattering and by high energy nucleon scattering experiments.

The author wishes to thank Associate Professor C. B. O. Mohr and Mr. B. A. Robson for helpful discussions on this work. The author is indebted also to the Commonwealth Scientific and Industrial Research Organization for an Australian Studentship.

References

Fregeau, J. H., and Hofstadter, R. (1955).—*Phys. Rev.* **99** : 1503.

Gatha, K. M., Patel, N. J., and Patel, P. F. (1954).—*Proc. Phys. Soc. Lond.* A **67** : 1111.

Gatha, K. M., Shah, G. Z., and Patel, N. J. (1954).—*Proc. Phys. Soc. Lond.* A **67** : 773.

Hofstadter, R., Fechter, H. R., and McIntyre, J. A. (1953).—*Phys. Rev.* **92** : 978.

Mohr, C. B. O., and Robson, B. A. (1956).—*Proc. Phys. Soc. Lond.* A **69** : 365.

Richardson, R. E., Ball, W. P., Leith, C. E., and Moyer, B. J. (1952).—*Phys. Rev.* **86** : 29.

Tassie, L. J. (1956).—*Proc. Phys. Soc. Lond.* A **69** : 205.

Williams, R. W. (1955).—*Phys. Rev.* **98** : 1387.

THE THERMAL EXPANSION OF A GROSSULARITE GARNET*

By L. G. Wilson†

Introduction

A study of the thermal behaviour of the lattice constant of a grossularite garnet in terms of thermal expansion has been made using a 10·302 cm diameter high temperature back reflection symmetrical processing X-ray camera designed by G. B. Clarke (1951) for the measurement of mineral lattice constants at high temperature.

The measurement of the diffraction patterns obtained in the work was made with a thin steel scale graduated in 0·5 mm. A reproducibility of ±0·01 cm. was obtained in these measurements. This method was adopted in preference to that employing a travelling microscope with unit magnification, which gave a similar reproducibility of measurement. The film measurements were converted to lattice constants using the extrapolation technique of Cohen (1935). $K\alpha_1$ and $K\alpha_2$ reflections always appeared resolved and, since unfiltered radiation was used, $K\beta_1$ reflections frequently occurred. For calculation all reflections observed were reduced to equivalent $K\alpha_1$ reflections.

The behaviour of the camera was tested by making a series of measurements, at various temperatures, of the lattice constant of platinum and comparing the equation relating these values to their corresponding temperatures, with accepted lattice constant-temperature expressions for platinum. The results of this work are given below.

Platinum Lattice Constant Test

Twenty-seven lattice constant determinations were made on a sample of platinum at different temperatures over the range 15–800 °C using cobalt radiation ($K\alpha_1=1\cdot78529$ kX). Using the least squares method the equation

---

* Manuscript received May 16, 1956.

† Physics Department, University of Western Australia, Perth; present address: C.S.I.R.O., East Melbourne.

best relating these lattice constants a to temperature t °C for the experimental values was calculated as

$$a=3\cdot91414+3\cdot80\times10^{-5}t\ (\mathrm{kX}), \quad \ldots\ldots\ldots\ldots \quad (1)$$

where the standard error of a is $\pm3\times10^{-4}$ and those for the constants of the equation are $\pm7\times10^{-5}$ and $\pm2\times10^{-7}$ respectively.

A reasonable frequency distribution is shown by the deviations of the values of a, calculated from equation (**1**) for the experimental temperatures, from the experimental values, only one value exceeding twice the standard error.

The validity of equation (**1**) as a means of expressing platinum lattice constants at different temperatures was tested by comparing the value of a calculated at 100 °C intervals using equation (**1**) and the values obtained from two expressions given in the International Critical Tables (1927) together with the lattice constant value of 3·91519 kX at 25 °C given by the Alphabetical and Numerical Indexes of X-ray Diffraction Patterns (American Society for Testing Materials 1953). The two expressions used are

$$a=a_0(1+8\cdot786\times10^{-6}t+3\cdot118\times10^{-9}t^2 +5\cdot2\times10^{-12}t^3+4\cdot095\times10^{-15}t^4), \quad \ldots\ldots\ldots \quad (2)$$

and

$$a=a_0(1+9\cdot75\times10^{-6}t). \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

Equation (**2**) is given as valid over the temperature range from −50 to +600 °C and equation (**3**) over the range from 0 to 1670 °C.

The comparison shows that equation (**1**) expresses the lattice constant of platinum to an agreement of $\pm2\times10^{-4}$ kX with other available data, and therefore the apparatus can be regarded as a satisfactory instrument for lattice constant determinations over the range from 15 to 800 °C.

The Thermal Expansion of Grossularite Garnet

The grossularite garnet used for these determinations was originally supplied by the Foote Mineral Company who suggest its origin as Boleo, near Santa Rosalie, Lower California, Mexico. An analysis indicated the specimen to be a normal grossularite garnet and an optical examination shows the material to be uniform in composition. Pure $3CaO.Al_2O_3.3SiO_2$ comprised 93·33 per cent. of the material and iron oxides amounted to 5·61 per cent.

Using iron radiation ($K\alpha_1=1\cdot93207$ kX) 28 lattice constant determinations were made with the specimen held at different temperatures over the range 25–940 °C. The equation best relating these values of lattice constant and temperature estimated by the least squares method is

$$a=11\cdot8368+8\cdot6\times10^{-5}t+2\cdot2\times10^{-8}t^2\ (\mathrm{kX}), \quad \ldots\ldots\ldots\ldots \quad (4)$$

where a has a standard error of $\pm10^{-3}$ kX and the three constants of the equation have standard errors of $\pm4\times10^{-4}$, $\pm6\times10^{-6}$, and $\pm3\times10^{-9}$ respectively.

As in the case of the platinum results a reasonable frequency distribution is shown by the deviations of the values of a, calculated from equation (**4**) for the experimental temperatures, from the experimental values, only two deviations exceeding twice the standard error.

Equation (4) when differentiated with respect to temperature gives an expression for the coefficient of thermal expansion of grossularite garnet, thus

$$\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}=\frac{1}{a}(8\cdot 6\times 10^{-5}+4\cdot 4\times 10^{-8}t). \quad \ldots\ldots\ldots\ldots \quad (5)$$

This equation expresses the coefficient to within ±2 per cent. at 25 °C varying proportionately to ±6 per cent. at 940 °C as a result of the standard errors involved in the estimation of the constants in equation (4).

The author wishes to thank Professor C. J. B. Clews for permission to carry out this work at the Physics Department, University of Western Australia. Acknowledgment is also made to Mr. J. Shearer for his valuable advice and guidance.

References

American Society for Testing Materials (1953).—Alphabetical and Numerical Indexes of X-ray Diffraction Patterns. Card No. 4–0787.

Clarke, G. B. (1951).—The measurement of the thermal expansion of grossularite by X-rays. M.Sc. Thesis. University of Western Australia.

Cohen, M. U. (1935).—Precision lattice constants from X-ray powder photographs. *Rev. Sci. Instrum.* **6**: 68–74; **7**: 155.

International Critical Tables (1927).—Vol. 2, pp. 461–2. (McGraw-Hill: New York.)

A MODEL OF NUCLEAR SHAPE OSCILLATIONS FOR γ–TRANSITIONS AND ELECTRON EXCITATION

By L. J. Tassie*

[*Manuscript received August* 1, 1956]

Summary

A model of nuclear shape oscillations is set up for an arbitrary nuclear charge distribution. For a uniform charge distribution the model reduces to the liquid drop model. The model is used to consider γ-transitions and electron excitation of nuclei. Explicit expressions are obtained for four charge distributions: (*a*) uniform, (*b*) Gaussian, (*c*) exponential, (*d*) uniform with Gaussian " edge ". The theory predicts a relative angular distribution of electrons scattered by the $4 \cdot 43$ MeV level of ^{12}C in agreement with the experimental results of Fregeau and Hofstadter (1955), but gives a scattered intensity seven times too large.

I. Introduction

The liquid drop model (Fierz 1943; Bohr 1952; Jekeli 1952) of the nucleus involves the assumption that the nucleus has a well-defined surface and a uniform charge and mass density. Experiments on the elastic scattering of high energy electrons (Hofstadter, Fechter, and McIntyre 1953; Hofstadter *et al.* 1954) have demonstrated that these assumptions are not valid even for heavy nuclei. An attempt is made to obtain a modified hydrodynamical model which allows for non-uniform nuclear charge and mass density distributions. In the liquid drop model distortions of the surface of the nucleus are considered; in this model similar distortions are considered not of the surface but of the shape of the nucleus, the distortions having effect throughout the whole nucleus. Excited states of the nucleus are then due to oscillations of the shape of the nucleus, i.e. oscillations involving departures from spherical symmetry of the mass and charge distributions of the nucleus.

Longitudinal compressional waves may be expected in the nucleus, involving nodes in the radial density distributions. As in the liquid drop model (Bohr 1952), the energy of excitation of these modes of radial oscillation should be much greater than the energy of excitation of the shape oscillations. Thus, in treating low-lying nuclear energy levels, it is assumed that the radial dependence of the density distributions does not alter appreciably.

This shape oscillation model is then used to consider γ-transitions and electron excitation of nuclei. For a uniform charge distribution this model reduces to the usual liquid drop model.

---

* Physics Department, University of Melbourne; present address: Research School of Physical Sciences, Australian National University, Canberra.

II. FORMULATION OF THE MODEL

It is assumed that the nucleus is spherically symmetric in the equilibrium position. This confines our treatment to nuclei with spin 0 in the ground state. (To treat other nuclei similar modifications to those used to incorporate the liquid drop model in the collective model (Bohr 1952 ; Bohr and Mottelson 1953) would be necessary.) The charge and mass density distributions are

$$\left.\begin{array}{l} Ze\rho_0^C(r), \\ \text{and} \qquad Am\rho_0^M(r) \end{array}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

respectively, allowing for a possible difference in these two distributions. For convenience, the " size " of a distribution is defined as the root mean square radius S,

$$S=\int r^2\rho(r)\mathrm{d}V. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

From the analysis (Yennie, Ravenhall, and Wilson 1954 ; Brown and Elton 1955) of high energy electron scattering experiments (Hofstadter, Fechter, and McIntyre 1953 ; Hofstadter *et al.* 1954) S is given approximately by

$$S=(3/5)^{\frac{1}{2}}\times 1\cdot 20A^{\frac{1}{3}}\times 10^{-13}\text{ cm.} \qquad \ldots\ldots\ldots\ldots (3)$$

For the excited states of the nucleus, shape oscillations occur. The distortion throughout the nucleus is treated in the same way as the distortion of the nuclear surface in the liquid drop model. We assume that under distortion, an element of mass and charge moves from $\mathbf{r}_0$ to $\mathbf{r}$ *without alteration of the volume it occupies*, i.e.

$$\rho(\mathbf{r})=\rho_0(r_0). \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

This assumes that each element of mass and charge is incompressible, i.e. the nucleus is composed of an inhomogeneous incompressible fluid. The " shape " of the distortion can be given by

$$r-r_0=\sum_{l=2,\,m}\alpha_{lm}(r_0)\mathrm{Y}_{lm}(\theta,\,\varphi)r_0^{l-1}, \qquad \ldots\ldots\ldots\ldots (5)$$

where r and r_0 are as shown in Figure 1, and Y_{lm} is the spherical harmonic. As in the liquid drop model, it is assumed that the motion in the nucleus is irrotational. Then

$$\mathbf{v}=\nabla\Phi, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

where Φ is the velocity potential. We have

$$\partial\Phi/\partial r=(\partial r/\partial t)_\rho. \qquad \ldots\ldots\ldots\ldots\ldots\ldots (6a)$$

The equation of continuity

$$\nabla\cdot(\rho\mathbf{v})+\partial\rho/\partial t=0,$$

becomes

$$\nabla\rho\cdot\nabla\Phi+\rho\nabla^2\Phi+\partial\rho/\partial t=0.$$

We assume that the deformations of the nucleus are small, i.e. α_{lm} small, and consider all expressions only to the first non-vanishing term in α_{lm}. Then to the first order in α_{lm},

$$(\partial\rho/\partial r)(\partial\Phi/\partial r)+\rho\nabla^2\Phi+\partial\rho/\partial t=0.$$

Using equation (**6a**),

$$(\partial\rho/\partial r)_t(\partial\Phi/\partial r)_t=(\partial\rho/\partial r)_t(\partial r/\partial t)_\rho=-\partial\rho/\partial t,$$

and thus

$$\nabla^2\Phi=0. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

Then

$$\Phi(\mathbf{r})=\Sigma l^{-1}\beta_{lm}Y_{lm}(\theta,\ \varphi)r^l, \quad \ldots\ldots\ldots\ldots\ldots \quad (8)$$

and from equation (**6a**)

$$\beta_{lm}=\dot\alpha_{lm}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

and α_{lm} is independent of r.

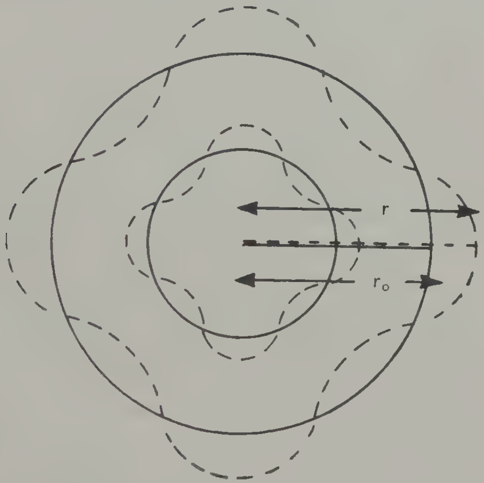


Fig. 1.—The shells of charge and mass in the equilibrium position of the nucleus, shown as continuous lines, move to the positions shown by the dotted lines in the distorted nucleus.

Since in the following, ρ will always appear multiplied by terms of the first order in α_{lm}, we need take $\rho(r)$ only to the zeroth order in α_{lm}, i.e. as $\rho_0(r)$ which we will now write as $\rho(r)$.

III. ELECTRON EXCITATION AND γ-TRANSITIONS

Using the first Born approximation, Schiff (1954) has developed expressions for the differential cross sections for the inelastic scattering of high energy electrons with excitation of various nuclear multipole transitions, and has shown that for collective modes of excitation only electric multipole transitions with $m=0$ need be considered. Then

$$d\sigma/d\omega=(d\sigma/d\omega)_P\,|\,\mathscr{F}_L\,|^2, \quad \ldots\ldots\ldots \quad (10)$$

where

$$(\mathrm{d}\sigma/\mathrm{d}\omega)_P=\tfrac{1}{4}Z^2(e^2/\hbar c)^2k^{-2}\cos^2\tfrac{1}{2}\theta\,\mathrm{cosec}^4\,\tfrac{1}{2}\theta \qquad \ldots\ldots (11)$$

is the point charge scattering cross section, and

$$\mathscr{F}_L=\{4\pi(2L+1)\}^{\frac{1}{2}}\int \mathrm{j}_L(qr)\mathrm{Y}_{L0}\rho_{\mathrm{trans}}\mathrm{d}V. \qquad \ldots\ldots (12)$$

$\mathrm{j}_L(z)\equiv(\pi/2z)^{\frac{1}{2}}\mathrm{J}_{L+\frac{1}{2}}(z)$ is the spherical Bessel function. $|\mathscr{F}_L|^2$ is called the scattering form factor. $\hbar k$ is the momentum of the incident electron, $q=2k\sin\frac{1}{2}\theta$, and $Ze\rho_{\mathrm{trans}}$ is the transition charge density of the nucleus. In equation (**10**) it is assumed that the energy of the transition is small compared with the energy of the incident electron, and that the angle of scattering θ is not small. Both these requirements are satisfied in experiments (Fregeau and Hofstadter 1955), as the inelastic scattering at low energies or small angles is masked by the bremsstrahlung tail of the elastic scattering.

From the equation of continuity,

$$\begin{aligned}\rho_{\mathrm{trans}}&=(\mathrm{i}\omega)^{-1}\nabla\cdot\mathbf{j}\\&=(\mathrm{i}\omega)^{-1}\nabla\cdot(\rho^C(r)\mathbf{v})\\&=(\mathrm{i}\omega)^{-1}\nabla\rho^C(r)\cdot\nabla\Phi\\&=(\mathrm{i}\omega)^{-1}(\mathrm{d}\rho^C/\mathrm{d}r)(\partial\Phi/\partial r), \qquad \ldots\ldots\ldots\ldots\ldots (13)\end{aligned}$$

using $\nabla^2\Phi=0$. $\hbar\omega$ is the energy of the transition. From equation (**8**) it follows that

$$\rho_{\mathrm{trans}}=(\mathrm{i}\omega)^{-1}\Sigma\beta_{lm}\mathrm{Y}_{lm}r^{l-1}\mathrm{d}\rho^C/\mathrm{d}r. \qquad \ldots\ldots\ldots\ldots (14)$$

From the orthogonality of the Y_{lm}, only the term $l=L$, $m=0$ will contribute to (**12**), giving

$$\begin{aligned}\mathscr{F}_L&=\{4\pi(2L+1)\}^{\frac{1}{2}}(\mathrm{i}\omega)^{-1}\beta_{L0}\int_0^\infty \mathrm{j}_L(qr)r^{L+1}(\mathrm{d}\rho^C/\mathrm{d}r)\mathrm{d}r\\&=\{4\pi(2L+1)\}^{\frac{1}{2}}(\mathrm{i}\beta_{L0}/\omega)q\int_0^\infty \mathrm{j}_{L-1}(qr)r^{L+1}\rho^C(r)\mathrm{d}r. \qquad \ldots\ldots (15)\end{aligned}$$

The transition probability for the emission of a photon of electric multipole order L, and energy $\hbar\omega$ is given by (Blatt and Weisskopf 1952)

$$T_E(L,M)=\frac{8\pi(L+1)}{L\{(2L+1)!!\}^2}\,\frac{(\omega/c)^{2L+1}}{\hbar}\,|Q_{LM}|^2, \qquad \ldots\ldots (16)$$

where

$$Q_{LM}=Ze\int r^L\mathrm{Y}_{LM}{}^*\rho_{\mathrm{trans}}\mathrm{d}V \qquad \ldots\ldots\ldots\ldots\ldots\ldots (17)$$

is the transition 2^L-pole moment. For the shape oscillation model, using equation (**14**), this becomes

$$\left.\begin{aligned}Q_{LM}&=(\mathrm{i}\beta_{LM}/\omega)Ze(2L+1)\int_0^\infty \rho^C(r)r^{2L}\mathrm{d}r\\&=(\mathrm{i}\beta_{LM}/\omega)Ze(2L+1)\overline{r_C^{2(L-1)}}/4\pi.\end{aligned}\right\} \qquad \ldots\ldots (18)$$

Comparing equations (**12**) and (**17**), electron excitation and γ-emission are related by

$$\lim_{q\to 0} \mathscr{F}_L=\{4\pi(2L+1)\}^{\frac{1}{2}}\frac{q^L}{(2L+1)!!}Q_{L0}/Ze. \quad \ldots\ldots\ldots \quad (\mathbf{19})$$

It is convenient to put

$$\mathscr{F}_L=S^{-L}(Q_{L0}/Ze)G_L(qS). \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{20})$$

Then the relative angular distribution of the inelastically scattered electrons depends on the function $G_L(qS)$, and the actual magnitude of the cross section is determined by the Q_{L0}. This gives us a method of determining the $|Q_{L0}|^2$ from electron scattering measurements by comparing the experimental form factor, i.e. the ratio of the actual scattering to point charge scattering, with G_L^2.

$$G_L(qS)=4\pi\{4\pi/(2L+1)\}^{\frac{1}{2}}(\overline{r_C^{2(L-1)}})^{-1}S^Lq$$
$$\times\int_0^\infty \mathrm{j}_{L-1}(qr)r^{L+1}\rho^C(r)\mathrm{d}r. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{21})$$

For all charge distributions,

$$\lim_{q\to 0} G_L=\{4\pi(2L+1)\}^{\frac{1}{2}}(qS)^L/(2L+1)!! \quad \ldots\ldots\ldots\ldots \quad (\mathbf{22})$$

In Section VI we obtain a theoretical estimate of Q_{L0}.

IV. Calculation of G_L

The inelastic scattering has been considered for four different charge distributions.

(*a*) Uniform distribution

$$\left.\begin{aligned}\rho(r)&=(3/4\pi)R_0^{-3}, \quad r\leqslant R_0,\\ \rho(r)&=0 \qquad\qquad\ \ , \quad r\geqslant R_0,\\ R_0&=(5/3)^{\frac{1}{2}}S.\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{23})$$

For the uniform distribution, the results are the same as those obtained from the usual liquid drop model.

$$G_L(qS)=\{4\pi(2L+1)\}^{\frac{1}{2}}(3/5)^{L/2}\mathrm{j}_L\{(5/3)^{\frac{1}{2}}qS\}. \quad \ldots\ldots \quad (\mathbf{24})$$

(*b*) Gaussian distribution

$$\left.\begin{aligned}\rho(r)&=(\pi^{\frac{1}{2}}g)^{-3}\exp\,(-r^2/g^2),\\ g&=(2/3)^{\frac{1}{2}}S.\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (\mathbf{25})$$

The integral in equation (**21**) can be obtained using equation 2, p. 393 of Watson (1944), the confluent hypergeometric function involved reducing to the exponential function, finally giving

$$G_L(qS)=\{4\pi(2L+1)\}^{\frac{1}{2}}\frac{(qS)^L}{(2L+1)!!}\exp\,\{-(qS)^2/6\}. \quad \ldots\ldots \quad (\mathbf{26})$$

(*c*) Exponential distribution

$$\left.\begin{aligned}\rho(r)&=(8\pi d^3)^{-1}\exp(-r/d),\\ d&=12^{-\frac{1}{2}}S.\end{aligned}\right\}\qquad\ldots\ldots\ldots\ldots\ (27)$$

For this case, the integral in equation (21) is obtained using equation 3, p. 393 of Watson (1944), the required hypergeometric function reducing to $(1-z)^{\frac{1}{2}}$, finally giving

$$G_L(qS)=\{4\pi(2L+1)\}^{\frac{1}{2}}\frac{(qS)^L}{(2L+1)!!}\{1+(qS)^2/12\}^{-(L+1)}.\quad\ldots.\ (28)$$

(*d*) Uniform distribution with a Gaussian " edge "

$$\left.\begin{aligned}\rho(r)&=\rho_0 && r\leqslant a,\\ \rho(r)&=\rho_0\exp\{-(r-a)^2/b^2\} && r\geqslant a,\\ \rho_0&=(3/4\pi)a^{-3}X(b/a)\\ S^2&=\frac{3}{5}a^2X(b/a)\Big\{1+\frac{5}{2}\pi^{\frac{1}{2}}b/a+10(b/a)^2+\frac{15}{2}\pi^{\frac{1}{2}}(b/a)^3\\ &\qquad+10(b/a)^4+\frac{15}{8}\pi^{\frac{1}{2}}(b/a)^5\Big\},\end{aligned}\right\}\quad\ldots\ (29)$$

where

$$X(b/a)=\Big\{1+\frac{3}{2}\pi^{\frac{1}{2}}b/a+3(b/a)^2+\frac{3}{4}\pi^{\frac{1}{2}}(b/a)^3\Big\}^{-1}.$$

$$\mathscr{F}_L=\{4\pi(2L+1)\}^{\frac{1}{2}}(\mathrm{i}\beta_{L0}/\omega)\rho_0a^{L+1}\{\mathrm{j}_L(qa)+qaf_L(b/a,\,qa)\},\quad\ldots\ (30)$$

where

$$f_L(x,\,y)=\int_0^\infty\exp\{-(t-1)^2/x^2\}t^{L+1}\mathrm{j}_{L-1}(yt)\mathrm{d}t.\quad\ldots\ldots\ldots\ldots\ (31)$$

For electric quadrupole transitions the $G_2^2(qS)$ are shown in Figure 2. The values for distribution (*d*) were obtained for $b/a=0\cdot4$ by graphical integration of $f_2(x,y)$. For $b/a=0\cdot4$, this distribution is a reasonable approximation to the smoothed uniform charge distribution, for which the calculated elastic scattering of high energy electrons by gold (Brown and Elton 1955) agrees with experiment (Hofstadter, Fechter, and McIntyre 1953 ; Hofstadter *et al.* 1954).

V. COMPARISON WITH EXPERIMENT FOR ^{12}C

Because of its statistical nature, this shape oscillation model should be valid only for medium and heavy nuclei. Also, the inelastic electron scattering has been considered using the Born approximation, which breaks down for scattering by heavy elements. However, Yennie, Wilson, and Ravenhall (1953) have shown that the Born approximation is not very far out for the elastic scattering by copper, except in the neighbourhood of diffraction minima. Thus, our results would be expected to apply to nuclei with atomic numbers about 60. Because of the large uncertainty in theoretical nuclear matrix elements, there is little justification for using more accurate methods than the Born approximation at present.

However, the only experimental results available to date are for beryllium (McIntyre, Hahn, and Hofstadter 1954) and carbon (Fregeau and Hofstadter 1955). Comparison has been made with the experimental results of Fregeau and Hofstadter (1955) for the excitation of the 4·43 MeV level of carbon,

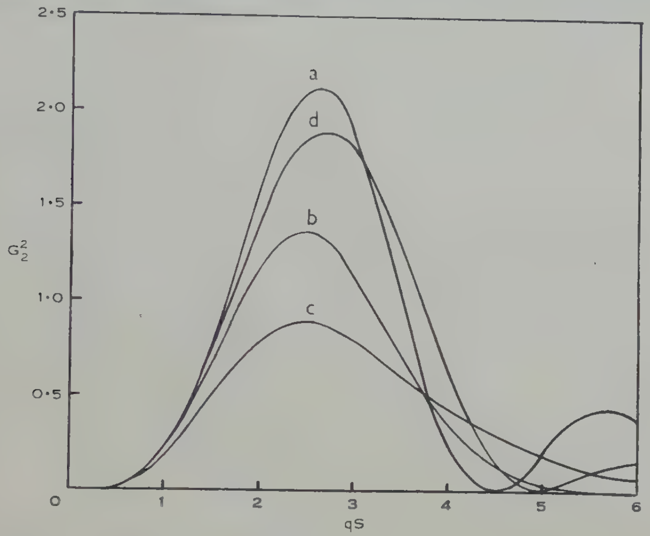


Fig. 2.—$\{G_2(qS)\}^2$ (see equation (**21**)) for electric quadrupole transitions for the following charge distributions : (*a*) uniform, (*b*) Gaussian, (*c*) exponential, (*d*) uniform with Gaussian " edge ", $b/a=0\cdot4$.

TABLE 1

EXCITATION OF THE 4·43 MeV LEVEL OF ^{12}C

Column 4 gives the results for $|Q_{20}|^2/e^2$ obtained by comparing the theoretical scattering form factors with the form factor obtained experimentally by Fregeau and Hofstadter (1955). Column 5 gives the theoretical value of $|Q_{20}|^2/e^2$ obtained from equation (**41**) in Section VI

Charge Distribution	S ($\times 10^{-13}$ cm)	Agreement of Theoretical with Experimental Form Factor	$\lvert Q_{20}\rvert^2/e^2$ From Comparison with Experimental Scattering Results ($\times 10^{-52}$ cm^4)	$\lvert Q_{20}\rvert^2/e^2$ From Theory of Section VI ($\times 10^{-52}$ cm^4)
(*a*) Uniform ..	2·20*	Good	7·7	55
	2·23	Good	7·9	56
(*b*) Gaussian ..	2·47*	Unsatisfactory	11·8	69
	2·04	Good	7·8	47
(*c*) Exponential ..	2·58*	No agreement		
	1·9	Good	10	41
(*d*) Uniform with Gaussian " edge " $b/a=0\cdot4$	2·4	Fair	10	65
	2·20	Good	8·4	55

* Indicates value of S giving best fit to the elastic scattering form factor for that particular charge distribution.

although one hardly expects this nuclear model to be valid for such a light element. Fregeau and Hofstadter obtain the following r.m.s. radii from their elastic scattering results,

$$\text{Uniform distribution } S=2\cdot 20\times 10^{-13}\text{ cm}$$
$$\text{Gaussian distribution } S=2\cdot 47\times 10^{-13}\text{ cm.}$$

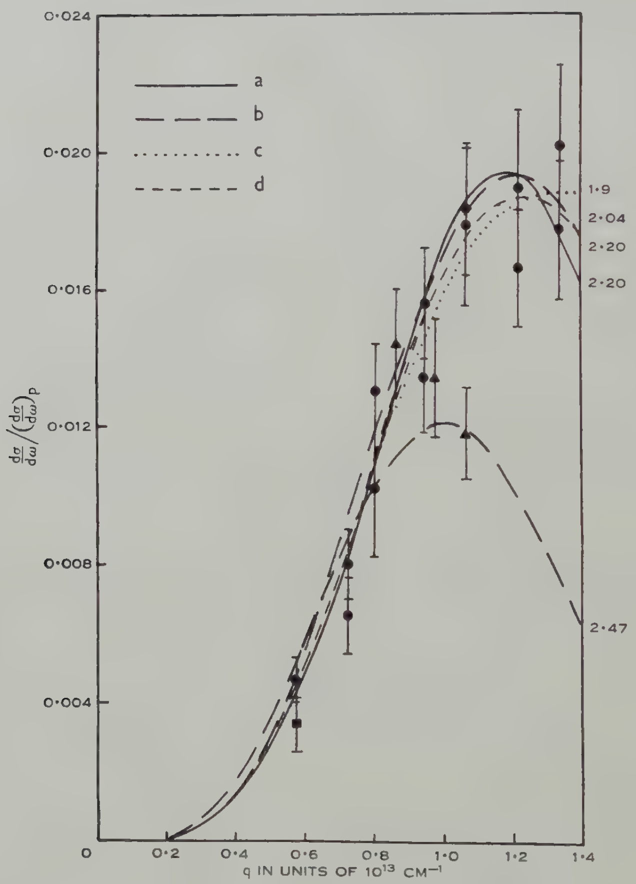


Fig. 3.—Ratio of cross section for the quadrupole excitation of the $4\cdot 43$ MeV level to the point charge scattering cross section of ^{12}C. Theoretical curves are $(S^2Ze)^{-2}\,|\,Q_{20}\,|^2\{G_2(qS)\}^2$ for the charge distributions: (*a*) uniform, (*b*) Gaussian, (*c*) exponential, (*d*) uniform with Gaussian " edge ", $b/a=0\cdot 4$. The curves are labelled with the value of S in units of 10^{-13} cm, and the $|\,Q_{20}\,|^2$ are given in column 4 of Table 1. The experimental points are those of Fregeau and Hofstadter (1955) for the scattering of electrons with energies 187 MeV (full circles), 150 MeV (triangles), and 80 MeV (square).

For an exponential charge distribution, the best fit to these elastic scattering results is obtained with $S=2\cdot 58\times 10^{-13}$ cm, but the fit is not satisfactory as the theoretical scattering cross section falls off too slowly with angle.

The $4\cdot 43$ MeV transition is electric quadrupole 0^+-2^+. The results obtained for $|\,Q_{20}\,|^2$ by comparing our $G_2^2(qS)$ using the above values of S, with the experimental form factor given by Fregeau and Hofstadter are given in Table 1, and the results for the form factor $(S^2Ze)^{-2}\,|\,Q_{20}\,|^2G_2^2$ are shown in Figure 3 in comparison with the experimental form factor. The experimental form factor could not be fitted using an exponential distribution with $S=2\cdot58\times10^{-13}$ cm.

The experimental results were then fitted by varying both S and $|\,Q_{20}\,|^2$ for all the charge distributions in Section IV including (*d*) with $b/a=0\cdot4$. The results are also shown in Figure 3 and Table 1. Under these conditions, the scattering form factors derived from all the distributions will fit the experimental form factor. However, only distributions (*a*) and (*d*) give a result for S consistent with the elastic scattering experiments, and the exponential distribution can be definitely excluded. The model used here is the same as the liquid drop model for the uniform distribution and close to it for distribution (*d*). The difference in S required to fit the Gaussian distribution to the elastic scattering and inelastic scattering may be due to the inadequacy of our model for this distribution. The experimental form factor can also be fitted by the form factor derived from a uniform transition charge density with a r.m.s. radius $S=2\cdot58\times10^{-13}$ cm, but it would seem that this transition charge density can be excluded since it has an appreciably larger extension than the static charge distribution. For the range of qS of these experiments, approximately $qS=1$ to $qS=3$, it seems that the scattering is largely model independent.

Experiments at larger values of qS, i.e. heavier nuclei, larger scattering angles, and higher electron energies, would give more information about the transition charge density and would provide a more critical test of the theory. However, we conclude from this analysis that the transition charge density has a larger r.m.s. radius than the static charge density and is greatest near the edge of the nucleus.

VI. THE TRANSITION MULTIPOLE MOMENTS

To obtain numerical values of the multipole moments from equation (**18**), we need the expectation value of β_{LM} for the transition. Using the quantum mechanical relation

$$\langle n\,|\,\dot{\alpha}_{lm}\,|\,n'\rangle=\mathrm{i}\omega\langle n\,|\,\alpha_{lm}\,|\,n'\rangle, \quad \ldots\ldots\ldots\ldots\ (\mathbf{32})$$

we obtain, in a similar way to the quantum treatment of the liquid drop model for small deformations (Bohr 1952), that for transitions from a no phonon to a one phonon state

$$\begin{aligned}\langle 0\,|\,\beta_{lm}\,|\,1\rangle&=\mathrm{i}\omega_l(\hbar\omega_l/2C_l)^{\frac{1}{2}}\\&=\mathrm{i}(\hbar\omega_l/2B_l)^{\frac{1}{2}},\end{aligned} \quad \ldots\ldots\ldots\ldots\ldots\ (\mathbf{33})$$

where the total energy of the shape oscillations of the nucleus is

$$\left.\begin{aligned}H&=T+V,\\T&=\tfrac{1}{2}\sum_{l=2,m}B_l\,|\,\beta_{lm}\,|^2,\\V&=\tfrac{1}{2}\sum_{l=2,m}C_l\,|\,\alpha_{lm}\,|^2.\end{aligned}\right\} \quad \ldots\ldots\ (\mathbf{34})$$

For this model it is easier to obtain an estimate of B_l than of C_l.

The total kinetic energy of the nuclear motion is

$$T=\tfrac{1}{2}Am\int\rho^M(r)v^2\mathrm{d}V. \tag{35}$$

Using equations (6) and (7), this reduces to

$$\begin{aligned} T&=-\tfrac{1}{2}Am\int\nabla\rho^M(r)\,.\,\Phi\nabla\Phi\mathrm{d}V\\ &=-\tfrac{1}{2}Am\int\Phi(\mathrm{d}\rho^M/\mathrm{d}r)(\partial\Phi/\partial r)\mathrm{d}V\\ &=-\tfrac{1}{2}Am\sum_{l=2,\,m}l^{-1}\,|\,\beta_{lm}\,|^2\int_0^\infty r^{2l+1}(\mathrm{d}\rho^M/\mathrm{d}r)\mathrm{d}r, \end{aligned} \tag{36}$$

using (Bohr 1952)

$$\beta_{lm}=(-1)^m\beta_{lm}{}^*. \tag{37}$$

This gives

$$\left.\begin{aligned} B_l&=l^{-1}(2l+1)Am\int_0^\infty r^{2l}\rho^M(r)\mathrm{d}r\\ &=l^{-1}(2l+1)(Am/4\pi)\overline{r_M^{2(l-1)}}. \end{aligned}\right\} \tag{38}$$

From equations (18), (33), and (38), we obtain for transitions from a one phonon state to a no phonon state or vice versa,

$$|\,Q_{LM}\,|^2=Z^2e^2\hbar(8\pi Am\omega_L)^{-1}L(2L+1)(\overline{r_C^{2(L-1)}})^2/\overline{r_M^{2(L-1)}}. \tag{39}$$

There has been some doubt whether the nuclear charge distribution is the same as the mass distribution. Purely nuclear measurements (Blatt and Weisskopf 1952, p. 15) have indicated a larger nuclear size than the electromagnetic measurements (Hofstadter, Fechter, and McIntyre 1953 ; Hofstadter *et al.* 1954). However, an analysis by Williams (1955) of the experiments of Coor *et al.* (1955) on the absorption cross sections of nuclei for 1·4 BeV neutrons yields a r.m.s. radius in agreement with that determined from electromagnetic measurements. Assuming then, that

$$\rho^C(r)=\rho^M(r)=\rho(r),$$

we obtain

$$|\,Q_{Lm}\,|^2=Z^2e^2\hbar(8\pi\omega_L Am)^{-1}L(2L+1)\overline{r^{2(L-1)}}. \tag{40}$$

If $\rho^C(r)$ and $\rho^M(r)$ are of the same form but with different S, then the correction factor to the left-hand side of equation (40) would be $(S_C/S_M)^{2(L-1)}$.

For the distributions in Section IV, $\overline{r^{2(L-1)}}$ is given by

(*a*)	Uniform	$3(2L+1)^{-1}(5/3)^{L-1}S^{2(L-1)}$,
(*b*)	Gaussian	$3^{1-L}(2L-1)!!S^{2(L-1)}$,
(*c*)	Exponential	$\tfrac{1}{2}\times 12^{1-L}(2L)!S^{2(L-1)}$.

For different distributions with the same r.m.s. radius, the difference in the $|\,Q_{LM}\,|^2$ are small for the first few L and increase with L.

For electric quadrupole transitions,

$$|Q_{2M}|^2=Z^2e^2\hbar(\pi\omega_L Am)^{-1}(5/4)S^2. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{41})$$

The values obtained from equation (**41**) are compared in Table 1 with the values derived from the inelastic electron scattering measurements for the various values of S. It is seen that the theoretical value is approximately seven times too large.

From the experimental results (Fregeau and Hofstadter 1955), we can deduce that $|Q_{20}|^2/e^2$ is approximately 8×10^{-52} cm^4. Inserting this value in equation (**16**), gives 6×10^{-14} sec as the lifetime of the 4·43 MeV level of ^{12}C for γ-decay to the ground state. This is consistent with the result of the γ-decay experiments of Mills and Mackin (1954) that the lifetime of this state is less than 3×10^{-13} sec. However, Devons, Manning, and Towle (1956) have measured the lifetime of this level for γ-emission using a recoil method and obtain a value of $(2\cdot6\pm0\cdot9)\times10^{-14}$ sec.

VII. Discussion

In conclusion, this theory of nuclear shape oscillations seems capable of explaining the angular distribution of the scattered electrons, but gives a scattered intensity too large by a factor of 7.

The assumption in equation (**4**) may cause some error because of compression effects, even when there are no compressional waves present. Woeste (1952) has treated compression effects for a density distribution with a sharp edge and only small deviations from uniformity, and shows that the density is greater at the surface than at the centre of the nucleus. However, electron scattering experiments (Hofstadter, Fechter, and McIntyre 1953 ; Hofstadter *et al.* 1954) show that the nuclear charge density distribution does not have a sharp surface, but that the density decreases smoothly to zero over a finite distance. This non-uniformity of the density distribution is more easily explained by quantum mechanics than by a classical effect such as compressibility. The nucleon wave functions must be smooth, and therefore must decrease smoothly to zero, giving rise to non-uniform nuclear charge and mass density distributions.

It is doubtful whether the similarity of the nucleus to an inhomogeneous fluid is such as to justify a detailed treatment including compression effects. A more accurate theory of the electron excitation of collective oscillations could be obtained by using the theory of an oscillating shell structure given by Araújo (1956). Such a theory would be very complicated for heavy nuclei, and would be restricted to the density distributions which can be obtained from simple nuclear potentials such as the spherical box and parabolic well.

At any rate, the theory developed here shows that the diffuseness of the nuclear surface can be neglected when treating the electron excitation of collective transitions provided that qS is not too large, as the curves for G_2^2 in Figure 2 for charge distributions (*a*), (*b*), and (*d*) do not differ very greatly up to about $qS=2$.

VIII. Acknowledgments

The author would like to thank Associate Professor C. B. O. Mohr for his interest in this work. He is indebted also to the Commonwealth Scientific and Industrial Research Organization for an Australian Studentship.

IX. References

Araújo, J. M. (1956).—*Nuclear Physics* **1**: 259.

Blatt, J. M., and Weisskopf, V. F. (1952).—"Theoretical Nuclear Physics." (Wiley: New York.)

Bohr, A. (1952).—*Math.-fys. Medd.* **26**: No. 14.

Bohr, A., and Mottelson, B. R. (1953).—*Math.-fys. Medd.* **27**: No. 16.

Brown, G. E., and Elton, L. R. B. (1955).—*Phil. Mag.* **46**: 164.

Coor, T., Hill, D. A., Hornyak, W. F., Smith, L. W., and Snow, G. (1955).—*Phys. Rev.* **98**: 1369.

Devons, S., Manning, G., and Towle, J. H. (1956).—*Proc. Phys. Soc.* A **69**: 173.

Fierz, M. (1943).—*Helv. Phys. Acta* **16**: 365.

Fregeau, J. H., and Hofstadter, R. (1955).—*Phys. Rev.* **99**: 1503.

Hofstadter, R., Fechter, H. R., and McIntyre, J. A. (1953).—*Phys. Rev.* **92**: 978.

Hofstadter, R., Hahn, B., Knudsen, A. W., and McIntyre, J. A. (1954).—*Phys. Rev.* **95**: 512.

Jekeli, W. (1952).—*Z. Phys.* **131**: 481.

McIntyre, J. A., Hahn, B., and Hofstadter, R. (1954).—*Phys. Rev.* **94**: 512.

Mills, W. R., and Mackin, R. J. (1954).—*Phys. Rev.* **95**: 1206.

Schiff, L. I. (1954).—*Phys. Rev.* **96**: 765.

Watson, G. N. (1944).—"Theory of Bessel Functions". 2nd Ed. (Cambridge Univ. Press.)

Williams, R. W. (1955).—*Phys. Rev.* **98**: 1387.

Woeste, K. (1952).—*Z. Phys.* **133**: 370.

Yennie, D. R., Ravenhall, D. G., and Wilson, R. N. (1954).—*Phys. Rev.* **95**: 500.

Yennie, D. R., Wilson, R. N., and Ravenhall, D. G. (1953).—*Phys. Rev.* **92**: 1325.

BOUNDS FOR THE COLLAPSE LOAD OF A BEAM COMPRESSED BY THREE DIES

By W. Freiberger*

[*Manuscript received February* 13, 1956]

Summary

This paper deals with the problem of the plastic deformation of a beam under the action of three perfectly rough rigid dies, two dies applied to one side, one die to the other side of the beam, the single die being situated between the two others. It is treated as a problem of plane plastic flow. Discontinuous stress and velocity fields are assumed and upper and lower bounds for the pressure sufficient to cause pronounced plastic yielding determined by limit analysis.

The treatment is approximate and subject to experimental verification.

I. Introduction

This problem had its origin in an enquiry regarding the straightening of the spar-beams for an aircraft. The beams were received bent and it was required to straighten them without introducing severe tensile stresses. It was suggested that by applying pressure successively along the beam by three rigid dies, two on one, one on the opposite side of the beam (Fig. 1), it could be straightened by introducing, mainly, shear deformation.

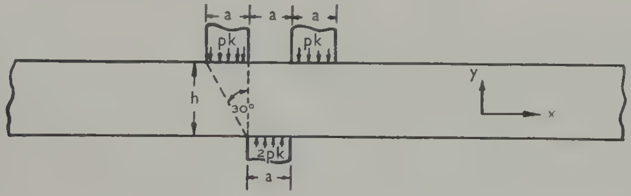


Fig. 1.—Beam under the action of three rigid, perfectly rough dies.

The present paper deals with this problem in a highly idealized form. The approximate solution presented here enables an estimate to be formed of the die pressure necessary to cause overall plastic deformation and also suggests a picture of the change of shape of the boundary when the dies are pressed together with given velocity.

II. General Theory

The solution of a problem in plane plastic flow consists of the determination of the three stress components σ_x, σ_y, and τ_{xy}, and the two velocity components u and v. The first and second slip lines are defined as lines in the directions

---

* Aeronautical Research Laboratories, Department of Supply, Melbourne ; present address : Brown University, Providence, R.I., U.S.A.

(the first and second shear direction) obtained by 45° counterclockwise rotations of the directions (first and second principal direction) across which the tensile stress is a maximum and minimum, respectively. These lines are orthogonal to each other and it can be shown (Prager and Hodge 1951) that

$$\left.\begin{aligned}\omega-\theta&=\text{constant along a first slip line,}\\ \omega+\theta&=\text{constant along a second slip line,}\end{aligned}\right\} \qquad \ldots\ldots\ldots \quad (1)$$

where $2k\omega$ is the mean normal stress (k is the yield stress in pure shear and so $2k$ the yield stress in pure tension), and θ is the angle between the first slip line and the negative y-axis. In terms of ω and θ, the stress components are given by

$$\left.\begin{aligned}\sigma_x&=2k\omega+k\sin 2\theta,\\ \sigma_y&=2k\omega-k\sin 2\theta,\\ \tau_{xy}&=-k\cos 2\theta.\end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

These satisfy the equilibrium equations and the yield condition

$$(\sigma_x-\sigma_y)^2+4\tau_{xy}^2=4k^2.$$

It has been shown (Prager 1948; Freiberger 1956) that in cases where it is impossible to satisfy all the boundary conditions, discontinuities may be introduced to give a solution. Moreover, in some cases which admit continuous solutions, it is possible to construct discontinuous solutions as a device to obtain approximate solutions more easily (Hodge 1950; Prager and Hodge 1951). If a line of discontinuity (" shock ") separates two regions of perfectly plastic material, then the following relation must be valid (Prager 1948):

$$\left.\begin{aligned}\theta_1+\theta_2&=2\alpha_{12},\\ \omega_2-\omega_1&=\sin(\theta_1-\theta_2).\end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

Here, subscripts 1 and 2 refer to opposite sides of the shock, α_{12} is the angle between the shock and the negative y-axis. The first equation (**1**) means that the shock must bisect the intersecting slip lines of either family. Conservation of mass requires the velocity component normal to the shock to be continuous across the shock.

To estimate the load at which sufficient plastic flow has taken place in the beam for pronounced overall yield to take place (" collapse ") we make use of the theorems of limit analysis (Hodge 1950; Prager and Hodge 1951) which are true for continuous and discontinuous fields of stress and velocity. These theorems are:

Theorem 1.—Collapse will not occur until the largest values of the surface tractions are reached for which it is possible to find a statically admissible stress field.

Theorem 2.—Collapse will occur under the smallest values of the surface tractions for which it is possible to find a kinematically admissible velocity field.

A stress field is called statically admissible if it satisfies the equations of equilibrium and the boundary conditions and nowhere violates the yield condition.

A velocity field is called kinematically admissible if it satisfies the incompressibility condition and the boundary conditions and if the rate at which the applied tractions do work on the velocities of their points of application equals (or is greater than) the rate of internal energy dissipation. For a surface of discontinuity, the rate of internal energy dissipation per unit area at any point of the surface is (Shield and Drucker 1953)

$$D=kv, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

where v is the change in the tangential velocity component across the surface of discontinuity at the point.

III. Velocity Field—Upper Bound for Limit Pressure

The slip line field of the right-hand side of Figure 2 will be assumed for the velocity distribution; the surfaces of the dies being perfectly rough, the slip lines meet the dies at right angles. It is also seen that the first of equations (3) is satisfied across each shock.

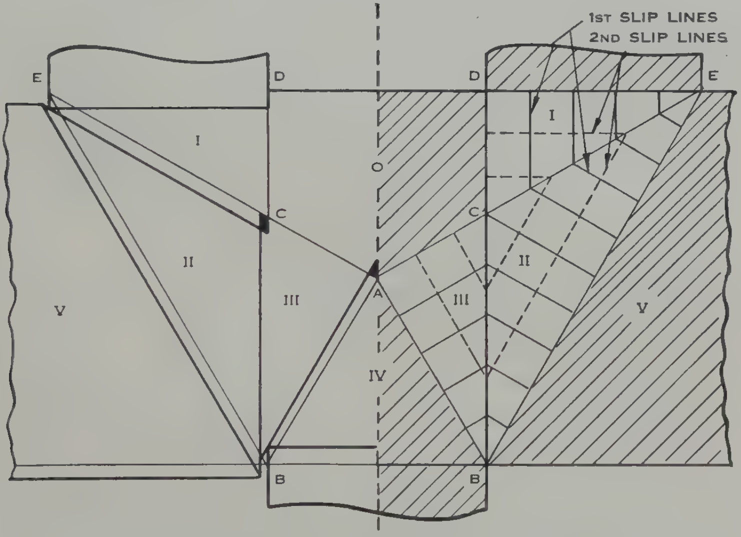


Fig. 2.—Left-hand side: boundaries after incipient plastic flow. Right-hand side: slip line field at incipient plastic flow, shaded areas rigid.

Let the pressure per unit length on each of the upper dies be pk; for the lower die it is therefore $2pk$, by equilibrium. Since the assumed relative velocity between upper and lower dies is irrelevant, let the dies move with unit velocity, the upper dies downwards, the lower die upwards.

Let U and V be the velocities along first and second slip lines, respectively, and denote the region to which they refer by a subscript 1, 2, or 3. With reference to Figure 2, applying the condition that normal velocity components must be continuous across lines of discontinuity and applying the boundary condition that dies move with unit velocity, we obtain :

$$\begin{aligned}
&\text{across } EC : U_1 \cos 30 + V_1 \sin 30 = U_2 \cos 30 + V_2 \sin 30,\\
&\text{across } CB : U_3 \cos 30 - V_3 \sin 30 = U_2 \cos 30 - V_2 \sin 30,\\
&\text{across } AB : U_3 = v_4 \sin 30 = 1 \sin 30,\\
&\text{across } CA : V_3 = 0, \text{ across } DE : U_1 = 1, \text{ across } DC : V_1 = 0.
\end{aligned}$$

It follows that $U_1=1,\ U_2=1,\ U_3=\frac{1}{2},$
$V_1=0,\ V_2=\frac{1}{2}\sqrt{3},\ V_3=0,$

or, in terms of velocities along coordinate directions, with subscripts referring to regions :

$$\begin{aligned}
&u_0=0,\ u_1=0,\ u_2=-\tfrac{1}{4}\sqrt{3},\ u\ =-\tfrac{1}{4}\sqrt{3},\ u_4=0,\ u_5=u_2=-\tfrac{1}{4}\sqrt{3},\\
&v_0=0,\ v_1=-1,\ v_2=\tfrac{3}{4},\qquad v_3=\tfrac{1}{4},\qquad\quad v_4=1,\ v_5=-v_2=-\tfrac{3}{4}.
\end{aligned}$$

These values for the velocities are used in the left-hand side of Figure 3 to indicate the deformation of the boundaries at the moment of pronounced plastic flow. For this velocity field, clearly, $\theta_1=180°$, $\theta_2=60°$, $\theta_3=-60°$.

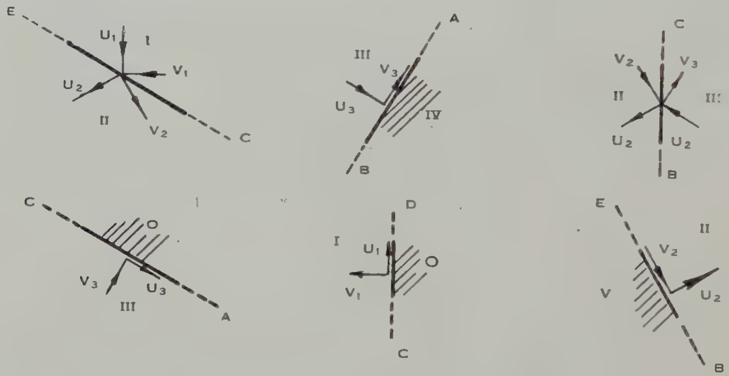


Fig. 3.—Conditions at boundaries of plastic regions.

There will be a value of $p=p^*$ such that if $p \geqslant p^*$ the velocity field of Figure 3 is a kinematically admissible velocity field and will yield an upper bound for the limit pressure. Denoting by $\overline{AB}$ etc., the lengths of the boundaries AB etc., and by v_{AB} etc., the change in the tangential velocity components across the lines of discontinuity, we have, for the rate of internal energy dissipation,

$$\begin{aligned}
D&=2k(\overline{AB}\,.\,v_{AB}+\overline{AC}\,.\,v_{AC}+\overline{BC}\,.\,v_{BC}+\overline{CD}\,.\,v_{CD}+\overline{EC}\,.\,v_{EC})\\
&=2k\left(1\ \cdot\ \frac{\sqrt{3}}{2}+\frac{1}{\sqrt{3}}\ \cdot\ \frac{1}{2}+\frac{2}{\sqrt{3}}\ \cdot\ 1+\frac{2}{\sqrt{3}}\ \cdot\ \frac{1}{2}\right)=4(\sqrt{3})k. \quad . (5)
\end{aligned}$$

The rate of external work done by the dies is

$$W=4apk. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Equating (5) and (6) it is found that the rate of external work done will not be less than the rate of dissipation of internal energy if the value of p is not less than

$$p^* = \sqrt{3}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

which is an upper bound for the limit pressure, by theorem 2.

The same upper bound is found by assuming the following displacement mode: the single die doubly shears a section of the beam and pushes it a small distance between the dies.[†]

IV. Stress Field—Lower Bound for Limit Pressure

To find, with the help of theorem 1, a lower bound for the limit pressure, a statically admissible stress field must be found. This can be done by using a discontinuous stress field for a truncated wedge (see Winzer and Carrier 1948), employed by Shield and Drucker (1953). This pattern is reproduced in Figure 4 (*a*) and the modified pattern for the present problem in Figure 4 (*b*). The latter is obtained from the former by superimposing a tension of value k in the region vertically below what is line DD in Figure 4 (*b*).

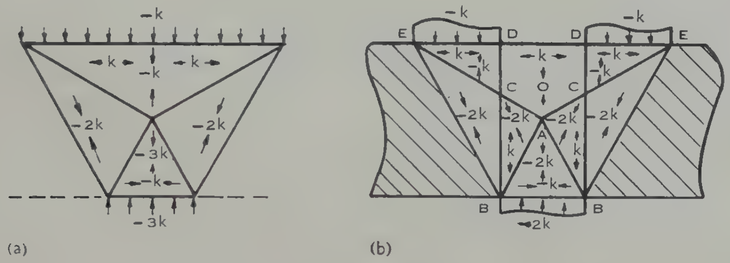


Fig. 4 (*a*).—Stress pattern for truncated wedge (Winzer and Carrier 1951).

Fig. 4 (*b*).—Modified stress pattern for beam.

Compression negative, tension positive.

This stress field, statically admissible for the present problem, at once gives for p the value

$$p^{**} = k \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

as a lower bound for the limit pressure on the upper dies.

Hence, combining the results of this section and those of Section III, we find, if P_L is the limit pressure,

$$k < P_L < k\sqrt{3}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

V. Conclusion

Under the assumption of perfect plasticity (flow under constant maximum shear stress) and plane strain, bounds are obtained for the collapse load of a beam subjected to compression by three rough rigid dies.

A discontinuous velocity field, satisfying velocity boundary conditions, is used to determine an upper bound to the limit pressure, i.e. the pressure per unit length on the upper dies sufficient to cause plastic flow throughout the cross section of the beam. This is found to have the value $k\sqrt{3}$. The velocity field is also used to give an indication of the deformation of the boundaries at the moment of pronounced plastic flow.

[†] This comment is due to Dr. J. M. Alexander, English Electric Company, England.

A discontinuous stress field, satisfying stress boundary conditions, is found to determine a lower bound for limit pressure. The value of this lower bound is k.

The difference between the upper and lower bounds is due to the different fields assumed for the velocity and stress solutions.

VI. Acknowledgment

The present paper was prepared as part of the research programme of the Aeronautical Research Laboratories, Melbourne, Vic., and is published by kind permission of the Chief Scientist, Department of Supply.

VII. References

Freiberger, W. (1956).—The elastic-plastic torsion of circular ring sectors. *Quart. Appl. Math.* **14**: 259.

Hodge, P. G. (1950).—Approximate solutions of problems of plane plastic flow. *J. Appl. Mech.* **17**: 257.

Prager, W. (1948).—" Discontinuous Solutions in the Theory of Plasticity." Courant Anniversary Vol. (Interscience Publishers: New York.)

Prager, W., and Hodge, D. G. (1951).—" Theory of Perfectly Plastic Solids." (John Wiley and Sons: New York.)

Shield, R. T., and Drucker, D. C. (1953).—The application of limit analysis to punch-indentation problems. *J. Appl. Mech.* **20**: 453.

Winzer, A., and Carrier, G. F. (1948).—The interaction of discontinuity surfaces in plastic fields of stress. *J. Appl. Mech., Trans. A.S.M.E.* **70**: 261.

CALCULATION OF ACTIVATION ENERGIES FROM CALORIMETRIC WORK CARRIED OUT AT A CONSTANT RATE OF HEATING

By J. F. Nicholas*

[*Manuscript received June 26, 1956*]

Summary

Some recovery processes in metals occur at temperatures too high to be readily accessible to isothermal calorimetric studies and such phenomena are more conveniently investigated using a constant rate of heating. It is shown that the activation energies of these processes can be determined from calorimetric work carried out at a constant rate of heating.

I. Introduction

In making calorimetric studies of the annealing of deformed metals, two principal methods have been adopted. The release of energy has been measured either at constant temperature or as the temperature is increased continuously. Both methods have inherent advantages. One feature of the isothermal technique is that the activation energies of the recovery processes can be readily determined (Gordon 1955). It is the purpose of this note to point out that activation energies can also be determined from calorimetric measurements of the type made by Clarebrough *et al.* (1952) and by Clarebrough, Hargreaves, and West (1955) where the rate of heating, i.e. the rate of increase of temperature, is constant. These experimental results comprise a series of curves, each of which shows the rate of release of energy from a deformed specimen as a function of temperature for a particular rate of heating. From the differences between the curves corresponding to different rates of heating, it is possible to deduce activation energies.

II. Method of Calculation

We assume that the state of the deformed metal can be described by one or more of a set of quantities such as the number of vacancies, the density of dislocations, the amount of metal recrystallized, etc. Then, over a limited range of temperature, we assume that the change in state can be ascribed effectively to a change in only one of these, denoted by x, and that the energy is released in proportion to the change in x. This implies that the total energy E which has been released at a given temperature T is a linear function of x.

We now make the usual assumption that the recovery process follows an equation of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t}=f(x)\mathrm{e}^{-Q/RT}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

---

* Division of Tribophysics, C.S.I.R.O., University of Melbourne.

where $f(x)$ is some function, independent of time and temperature, t is the time, Q is the activation energy of the process, and R is the gas constant. Then, since E is a linear function of x, this can be rewritten to give the rate of release of energy ΔP as

$$\Delta P=\frac{\mathrm{d}E}{\mathrm{d}t}=g(E)\mathrm{e}^{-Q/RT}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

where $g(E)$ is independent of time and temperature.

If the rate of heating is α, we can write

$$T=T_0+\alpha t, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

where T_0 is some initial temperature, and equation (2) can then be integrated formally to give

$$\begin{aligned}\int_0^E \frac{\mathrm{d}\xi}{g(\xi)}&=h(E),\ \text{say},\\ &=\frac{1}{\alpha}\int_{T_0}^{T} \mathrm{e}^{-Q/R\tau}\mathrm{d}\tau\\ &\sim\frac{1}{\alpha}\left[\frac{R\tau^2}{Q}\mathrm{e}^{-Q/R\tau}\left(1-2!\,\frac{R\tau}{Q}+3!\,\frac{R^2\tau^2}{Q^2}-\ldots\right)\right]_{T_0}^{T}, \quad \ldots\ldots \quad (4)\end{aligned}$$

the asymptotic expansion being derived by repeated integration by parts. Provided that T is sufficiently greater than T_0 and $RT \ll Q$, equation (4) can be written as

$$h(E)\simeq\frac{RT^2}{\alpha Q}\mathrm{e}^{-Q/RT}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

that is,

$$\ln\,(T^2/\alpha)\simeq Q/RT+\ln\,\{Qh(E)/R\}. \quad \ldots\ldots\ldots\ldots \quad (6)$$

If we now choose a particular value of E, say E_1, the last term of equation (6) is fixed and, for each heating rate, the value of T for this E_1 can be determined from the experimental results. Thus, if several heating rates have been used, Q may be estimated by plotting $\ln\,(T^2/\alpha)$ against $1/T$ and determining the slope of the resulting line. If only two heating rates have been used, Q is given by

$$Q=\frac{RT_1T_2}{T_1-T_2}\ \ln\ \frac{\alpha_1T_2^2}{\alpha_2T_1^2}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

The error made in neglecting all but the first term of the asymptotic expansion (4) can be corrected by a process of successive approximations.

This procedure is, in principle, capable of high accuracy since the mathematics can be made as accurate as is desired and several estimates for Q may be obtained by measuring values of T for several different values of E_1. However, in practice, a set of temperatures corresponding to a given E_1 is difficult to determine accurately because of uncertainties in the amount of energy released at low temperatures and in the exact position of the base level above which ΔP is measured. Such uncertainties, though small enough to be unimportant in the measurement of total energy evolved, may cause large errors in the estimation of Q since this depends essentially on the fairly small differences between the temperatures corresponding to a given E_1 for different rates of heating.

However, one set of temperatures which can be determined readily and accurately is the set of temperatures T^* at which the maxima in the curves of ΔP against T occur. Furthermore, it can be shown that these temperatures correspond to the release of approximately equal amounts of energy. For, since a maximum requires that $\mathrm{d}(\Delta P)/\mathrm{d}T=0$, we have, by equation (**2**),

$$\left\{g'(E)\frac{\mathrm{d}E}{\mathrm{d}T}+g(E)\frac{Q}{RT^2}\right\}\mathrm{e}^{-Q/RT}=0.$$

By using equations (**2**) and (**3**) to substitute for $\mathrm{d}E/\mathrm{d}T$ we get

$$g'(E)+(\alpha Q/RT^2)\mathrm{e}^{Q/RT}=0,$$

and hence, by use of equation (**5**),

$$g'(E)+1/h(E)\simeq 0. \quad \dots\dots\dots\dots\dots\dots \quad (8)$$

Since equation (**8**) is an equation in E only, its solution, which gives the value of energy for which ΔP is a maximum, is independent of the rate of heating. Thus, to the approximation implicit in equation (**5**), the temperatures T^* at which the maxima occur correspond to equal values of energy and these temperatures may be substituted into equations (**6**) or (**7**) to determine Q.

If the values of T^* are used to determine Q, the result cannot be corrected for the effect of cutting off the asymptotic expansion unless the form of $f(x)$ (and thus of $g(E)$ and $h(E)$) is known, either implicitly or explicitly. However, the effect of such corrections is not likely to be large and, in fact, in one case where equation (**6**) was used and $f(x)$ was known implicitly (Nicholas 1955), the correction to Q was less than 1 per cent., i.e. insignificant.

The advantages of using the values of T^* are that the temperatures can be taken directly from the experimental curves and will be insensitive to small errors in the position of the base line and to the amount of energy that is released at low temperatures. Furthermore, detailed consideration shows that the procedure will discriminate between two concurrently acting processes which have sufficiently different values of T^*. However, if one particular peak represents a two-stage process or a set of similar processes having almost identical activation energies, then equation (**6**) will only give some mean value of the energies involved and care must be taken in interpreting the result.

III. APPLICATIONS

Equation (**6**) can be used to determine activation energies for recrystallization from data given in Clarebrough, Hargreaves, and West (1955). Thus, for electrolytic copper (99·98 per cent. Cu) the activation energy for recrystallization after 45 per cent. elongation in tension is 33 kcal/mole, while for nickel (99·6 per cent. Ni) after heavy deformation in torsion, the corresponding value is 85 kcal/mole.

It is clear that the use of equation (**6**) need not be restricted to calorimetric studies even though these have the advantages that a quantity proportional to dx/dt is measured directly and that the amount of energy released by an imperfection is almost independent of the temperature of measurement. In fact, Parkins, Dienes, and Brown (1951) have already derived the equivalent of equation (**6**) and used it in resistivity studies during the " pulse-annealing " of AuCu. However, their derivation assumes that $f(x)$ is of the form x^{γ} and their integration of $1/f(x)$ is invalid for $\gamma=1$. Furthermore, they do not investigate the integral of exp $(-Q/RT)$ and assume that the variation, with temperature, of this integral is negligible compared to the variation of the integrand. Actually, their result is valid only because both the integral and the integrand vary with temperature in the same way.

IV. ACKNOWLEDGMENTS

It is a pleasure to acknowledge the continued cooperation of my colleagues L. M. Clarebrough, M. E. Hargreaves, and G. W. West.

V. REFERENCES

CLAREBROUGH, L. M., HARGREAVES, M. E., MICHELL, D., and WEST, G. W. (1952).—*Proc. Roy. Soc.* A **215** : 507.

CLAREBROUGH, L. M., HARGREAVES, M. E., and WEST, G. W. (1955).—*Proc. Roy. Soc.* A **232** : 252.

GORDON, P. (1955).—*J. Metals* **7** : 1043.

NICHOLAS, J. F. (1955).—*Phil. Mag.* **46** : 87.

PARKINS, W. E., DIENES, G. J., and BROWN, F. W. (1951).—*J. Appl. Phys.* **22** : 1012.

NEUTRON EXPERIMENTS WITH A SENSITIVE SZILARD–CHALMERS DETECTOR

By R. D. Edge*

[*Manuscript received July* 16, 1956]

Summary

The Szilard-Chalmers reaction has been developed as a sensitive detector for the comparison of neutrons from sources of different primary energy and angular distribution. Using sodium permanganate solution, a detection efficiency of approximately 7 per cent. for all neutrons in the energy range 0–5 MeV was obtained and high accuracy achieved.

The method of detection was employed to make the following measurements: (i) The neutron yield from the spontaneous fission of natural uranium was found to be $(1\cdot56\pm0\cdot12)\times10^{-2}$ neutrons $\sec^{-1}$ g^{-1}. (ii) The cross section for the reaction ^{6}Li(γ, n) at $6\cdot2$ MeV was found to be $(3\pm2)\times10^{-28}$ cm^2. (iii) The cross section for the reaction ^{9}Be(n, $2n$) induced by Ra-Be and Po-Be sources of neutrons was found to be $(0\cdot37\pm0\cdot06)\times10^{-24}$ cm^2. (iv) A limit was placed on the cross-over transition from the $2\cdot5$ MeV state to the ground state in ^{60}Co. The transition probability was found to be less than 1×10^{-6} of that for the cascade consisting of the two γ-rays.

I. Introduction

The Szilard-Chalmers reaction has been developed in the direction of the use of volumes as large as 300 gallons of aqueous solution, in order to extend the limit of detection of neutrons and allow the comparison of weak neutron sources (Edge 1956). In the case of the largest detector employed, less than 3 per cent. of neutrons having an energy of 10 MeV failed to be moderated and captured. As is well known, the Szilard-Chalmers reaction (1934) involves the decomposition of the molecule in which it occurs, as a result of the recoil following the emission of the prompt γ-rays emitted after neutron capture. The active component is thrown into another state of chemical combination which is capable of extraction from the remainder. The permanganate complex proved most suitable for the work, the activity being precipitated in the form of $^{56}MnO_2$ (Broda 1948).

When intense neutron sources were being measured, sufficient activity was induced in a saturated potassium permanganate solution, but for weak sources a sodium permanganate solution of 670 g/l concentration proved more sensitive.

II. Experimental Technique

A standard technique for processing the solution was developed in order to obtain reproducible results.

A proportion of the activity, called the " retentivity ", was never precipitated from solution. This proportion had to be kept constant, the two

---

* Research School of Physical Sciences, Australian National University, Canberra.

factors causing variation being temperature and pH. By maintaining the temperature to within ±2 °C, and holding the pH at (7 ±0·2) the variation in retentivity was found to be less than 2 per cent. In general, the pH of the solution remained constant and a buffer proved unnecessary, but it was necessary to keep the solution free of organic matter and to store it in darkness to prevent spontaneous precipitation of manganese dioxide.

Various methods of extracting the activity were investigated, such as centrifuging, electrostatic precipitation, filtration through asbestos wool, sintered glass, or filter paper. Only the last provided the reproducibility required, the most suitable paper tested being Whatman 540. It was found to be unnecessary to add a carrier to the solution to bring down the manganese dioxide ; however, a thin layer (1 g) of pure manganese dioxide powder spread evenly over the filter paper was required. To prevent the precipitate creeping round the edges of the paper in the Büchner funnel, a " Gaco " ring seal was pressed against the circumference of the filter paper by a glass chimney. Using such an arrangement it was found possible to vacuum filter 50 l. of solution in approximately 40 min.

After filtration the container was washed with distilled water and the washings themselves were filtered and collected separately ; less than 0·1 per cent. of the activity was left on the walls.

Before counting the precipitate was washed with distilled water followed by acetone and ether, to dry the paper. Counting was effected by placing the precipitate on the filter paper between two " Cellophane " sheets and wrapping this " sandwich " round a cylindrical thin-walled Geiger counter 3 cm in diameter and 10 cm long, with a wall thickness of only 0·007 in. With appropriate care it was found that the sandwich could be moved and then replaced accurately enough to give a counting rate reproducible to 2 per cent. The counter was fitted into a lead block to increase the backscattering of electrons and provide shielding against cosmic rays. Experiments on the self-absorption of the manganese dioxide indicated that slight variations in thickness of deposit over the area of the filter paper produced a negligible variation in the counting rate, provided the same quantity of manganese dioxide deposit was used each time.

III. Discussion of the Method of Detection

Four factors limited the usefulness of this detecting technique as a means of comparing weak neutron sources. These were the speed and efficiency with which the solution could be processed and the activity counted, the proportion of incident neutrons captured by manganese, as opposed to those that escaped or were captured by hydrogen, the long period activity present in the solution, and finally the cosmic ray background.

To facilitate rapid filtration a large filter paper, 9 cm in diameter, was used and the processing was reduced to the minimum number of steps compatible with maximum accuracy. It was found that, using a 50 l. tank, 90 min were required between the end of an irradiation and the commencement of counting.

Approximately 50 per cent. of the neutrons captured in the sodium permanganate solution were absorbed by ^{55}Mn and it was difficult to increase the fraction captured by manganese above this value as the solution became too concentrated.

After the decay of the 2·6 hr half-life (Hollander, Perlman, and Seaborg 1953) activity from ^{56}Mn, a considerable proportion of long period activity still remained (Dodson, Goldblatt, and Sullivan 1944) and this did not decay appreciably over the period of a year. In the case of potassium permanganate, it was thought to be caused by ^{40}K adsorbed on the filter paper; the much smaller effect in the case of sodium permanganate suggested that this may have been due to potassium impurities in the compound. In the experiments this long period activity was subtracted from the total activity in order to obtain the ^{56}Mn activity.

Although experiments were made intermittently over a period of 2 years, no significant temporal variation in cosmic ray neutron flux was detected. The mean of 10 determinations of the cosmic ray neutron flux at sea-level yielded a value of $2\cdot5\times10^{-3}$ neutrons cm^{-2} sec^{-1}, a value in reasonable agreement with other determinations (Kaplan and Yagoda 1952). This background was also taken into account in the experimental work with the detector.

The efficiency of a neutron detector is generally stated to be the number of counts recorded per neutron expressed as a percentage. In the case of a Szilard-Chalmers detector, however, this definition is difficult to interpret, since the decay of activity during and after irradiation requires that the sensitivity is a function not only of the length and variation of intensity during the irradiation but also of the period between the end of the irradiation and commencement of counting—hence, although up to 50 per cent. of the neutrons emitted by the source may be captured by the manganese in the solution, only a small proportion of the resulting activity would be recorded by the Geiger counter, part being lost in the extraction process, part by the decay, and part owing to the efficiency of the counting arrangement itself. A suitable definition of the sensitivity, which has been used by Dodson, Goldblatt, and Sullivan (1944), is the ratio expressed as a percentage of the activity registered by the counter, corrected to the end of the irradiation (long compared with the ^{56}Mn half-life), to the rate of emission of neutrons by the source. Using this definition the detection efficiency of the 50 l. tank of potassium permanganate was found to be 1 per cent., and of sodium permanganate 7 per cent., for a $RaBeF_4$ neutron source placed at the centre of the tank. The method, therefore, is the most sensitive yet applied, and compares with detectors such as " long " BF_3 counters or phosphors (Segre 1953).

The detection efficiency is energy dependent, being a function of the number of neutrons escaping the detector. Theoretical investigations of the fraction escaping were made using approximations such as the Fermi " age " method (Wallace and LeCaine 1947) (curve C of Fig. 1) and the two-group diffusion theory of Tait (1954) (curve B of Fig. 1). An experimental method was used to determine the number of neutrons of known energy escaping from the 50 l. sphere, and the energy-escape curve calculated using two-group diffusion theory was normalized using it. The resulting curve (A of Fig. 1) is considered to be accurate to 5 per cent. over the energy range 0–6 MeV.

The number of neutrons escaping was measured by comparing the relative activity induced by two neutron sources of known strength in two tanks, one of 50 and the other of 1134 l. (300 gal). The two sources used were a ^{24}Na D(γ, n) source, yielding neutrons of energy 0·26 MeV, and a RaBeF$_4$ source which gave neutrons of average energy 4·5 MeV, with a spread extending from low energies up to approximately 13 MeV. It was found that 73 per cent. of the RaBeF$_4$ neutrons and 98 per cent. of the ^{24}Na neutrons were captured in the 50 l. tank, the remainder escaping. At least 99 per cent. of the RaBeF$_4$ neutrons were captured in the large tank.

In the following section four experiments carried out using the technique and only possible with a sensitive neutron detector are described.

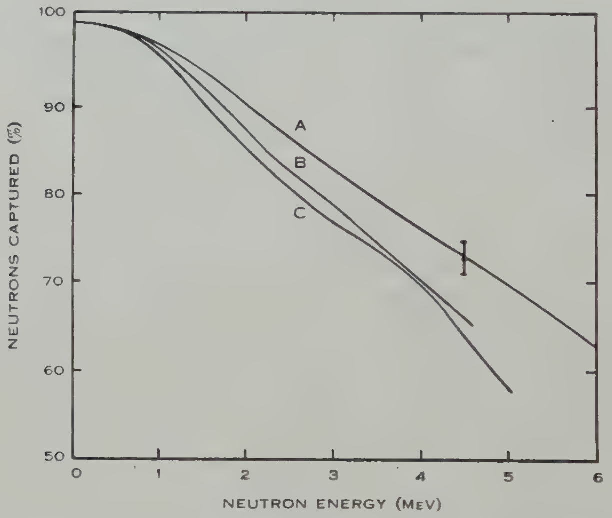


Fig. 1.—The variation of capture efficiency with energy for neutrons emitted from a source at the centre of a 50 l. sphere of potassium permanganate solution.
Curve *A*, the percentage of neutrons captured within the sphere, based on the two-group diffusion theory of Tait (1954) and normalized to pass through the experimental point shown, which was determined as described in the text; curve *B*, the percentage of neutrons captured according to the two-group diffusion theory of Tait (1954); curve *C*, the percentage of neutrons captured based on the diffusion theory given by Wallace and LeCaine (1947).

IV. Experimental Results

(*a*) *The Neutron Yield from the Spontaneous Fission of Natural Uranium*

Until 1952 the best value for the neutron yield from the spontaneous fission of natural uranium was that given by Segre (1952) as $(1{\cdot}5 \pm 0{\cdot}15) \times 10^{-2}$ neutrons sec^{-1} g^{-1} uranium. Littler (1952), however, has employed a subcritical pile to obtain $(1{\cdot}65 \pm 0{\cdot}1) \times 10^{-2}$ neutrons sec^{-1} g^{-1} uranium. In the present experiment an accurate comparison was made of the number of neutrons emitted from a block of uranium of known size with a RaBeF$_4$ neutron source, recently calibrated by Bretscher *et al.* (1949) and Edge (1956).

The cylindrical block of natural uranium metal (9·572 kg) was enclosed in a cadmium shell in order to reduce slow neutron fission, and was sealed in a glass container. The standard $RaBeF_4$ neutron source was similarly encased and also sealed in glass. Each, in turn, was then submerged at the centre of the spherical 50 l. glass tank of sodium permanganate solution (670 g/l), for a known period of several hours. On removal of a source the active $^{56}MnO_2$ deposit formed by the Szilard-Chalmers reaction was filtered off and counted as described above.

A sequence of three runs, alternating the uranium block and the standard source, was performed.

After subtracting the cosmic ray background and correcting for the escape of neutrons from the moderating medium, the ratio of $^{56}MnO_2$ activities gave the ratio of the number of neutrons emitted by the two sources, and from this the spontaneous fission rate of natural uranium was found to be $(1 \cdot 56 \pm 0 \cdot 12) \times 10^{-2}$ neutrons sec^{-1} g^{-1}.

(b) *The Cross Section for the Reaction* $^6Li\ (\gamma, n)$ *at* 6·2 *MeV*

A cylinder of lithium fluoride, 4 cm in diameter and 4 cm long, was irradiated by the γ-ray flux emitted from a thick calcium fluoride target, with which the cylinder was in contact, under bombardment by protons of 900 keV energy. The specially constructed target tube projected into the spherical 50 l. tank of sodium permanganate solution so that the source was located at the centre. The photoneutrons from the lithium which, from energy considerations, were the only possibilities, were captured in the tank and the active ^{56}Mn precipitated, extracted, and counted as before. For calibration the lithium was replaced by an identically cased cylinder containing heavy water, and in addition background runs were performed using a hollow cylinder. The cross section for $D(\gamma, n)$ at 6·2 MeV was taken from the work of Barnes *et al.* (1952) and from it the cross section for $^6Li(\gamma, n)$ at the same energy was found to be $(3 \pm 2) \times 10^{-28}$ cm^2. The rather large error in this case was due to the small isotopic abundance of ^6Li in natural lithium; separated ^6Li was not available in sufficient quantity for a better measurement. Titterton and Brinkley (1951) find the cross section has increased to $(5 \pm 2) \times 10^{-28}$ cm^2 at 17·6 MeV.

(c) *Cross Section for the Reaction* $^9Be(n, 2n)$

When beryllium is excited by incident neutrons of energy greater than 1·85 MeV the reaction $^9Be(n, 2n)^8Be$ is energetically possible and has been observed by Grace *et al.* (1951). The early investigations have produced conflicting results for the cross section (Edge 1956) and in order to try to resolve the discrepancies, three independent experimental methods have been employed.

In the first, neutrons produced by a Ra-Be source were used, and a comparison made of the neutrons scattered from beryllium with those from copper. The experiment was carried out as follows:

A collimated beam of neutrons was fired through a brass tube 4 cm in diameter which passed through the centre of a 10 l. spherical glass vessel filled with potassium permanganate solution. This was surrounded by a tank of borax solution, which acted as a neutron shield. Samples of beryllium or copper, in the form of cylinders 4 cm in diameter, were placed in the centre of this arrangement, so that the detector absorbed the neutrons scattered by these samples. After each exposure the activity was extracted from the solution and counted as before. A series of experiments were carried out and the $(n, 2n)$ cross section in beryllium was determined from the ratio of the relative average activities induced in the two samples.

Taking the values for the total neutron cross sections for beryllium and copper as given in the tables of Hughes and Harvey (1955) and correcting for the escape of neutrons from the detector, for scattered neutrons leading to the $(n, 2n)$ reaction, and for the (γ, n) reaction in the beryllium sample due to γ-radiation from the radium in the Ra-Be source, the cross section for Ra-Be neutrons was found to be $(0 \cdot 40 \pm 0 \cdot 06) \times 10^{-24}$ cm^2.

A second experiment was performed also with Ra-Be neutrons, employing a similar experimental arrangement, except that the tube ran only to the centre of the detector sphere. The total neutron flux was given by the induced activity with the beryllium block absent, the neutrons being stopped by the permanganate solution beyond the centre of the sphere. The additional induced activity with the block inserted was taken to be due to $(n, 2n)$ neutrons. Similar corrections were applied as in the previous experiment and the value $\sigma_{^9\mathrm{Be}(n,\,2n)} = (0 \cdot 386 \pm 0 \cdot 1) \times 10^{-24}$ cm^2 obtained.

For the third experiment a small Po-Be source was used, which had essentially the same neutron spectrum as the Ra-Be source. It was held between two beryllium cylinders, 4 cm in diameter and 4 cm long, and the whole was enclosed in a tight-fitting soft glass casing. The beryllium cylinders could be replaced by hollow cylinders of the same dimensions. This source, with and without the beryllium alternately, was placed for a known time at the centre of a spherical 50 l. glass vessel of potassium permanganate solution and the activities induced in the solution with and without the beryllium present determined. Errors and corrections were applied as before, except that in this case no correction for the (γ, n) reaction was necessary.

From the results the cross section for the $(n, 2n)$ reaction was found to be $\sigma_{^9\mathrm{Be}(n,\,2n)} = (0 \cdot 354 \pm 0 \cdot 08) \times 10^{-24}$ cm^2. The agreement between the three measurements is good and a combination of the three results allows a most probable value to be stated as $\sigma_{^9\mathrm{Be}(n,\,2n)} = (0 \cdot 37 \pm 0 \cdot 06) \times 10^{-24}$ cm^2. This cross section is much larger than is predicted by a theory based on the idea of an orbital neutron circulating around an ^{8}Be nucleus. Schlögl (1948) calculated a cross section of $0 \cdot 05 \times 10^{-24}$ cm^2 at an energy of $4 \cdot 5$ MeV using this model.

Experiments performed using neutrons from the $\mathrm{D}(d, n)$ reaction induced using a Cockcroft-Walton generator, gave some indication of a resonance in the neighbourhood of $2 \cdot 7$ MeV and this may be the same as that found by Stafford (1951) in the total cross section.

(*d*) *A Determination of the Cross-over Transition in ^{60}Co*

The γ-ray decay scheme of ^{60}Co is well known and consists of a transition from the $2 \cdot 50$ MeV state (spin and parity $4+$) to the $1 \cdot 33$ MeV state $(2+)$ and thence to the ground state $(0+)$. Myers and Wattenberg (1949) were able to place an experimental upper limit of 2×10^{-6} on the ratio of the number of transitions between the $2 \cdot 50$ MeV state and the ground state, to the number going to the $1 \cdot 33$ MeV state. The theoretical value for this transition E_3/E_2, based on a single particle model, according to these authors is 10^{-5} and a discrepancy exists which it is of interest to investigate.

An experiment was, therefore, performed in which two copper cylinders each containing 50 cm^3 of heavy water were supported so as to subtend the maximum possible solid angle at a source consisting of 25 curies of ^{60}Co. The E_2 transition γ-rays from the cobalt lie below the threshold for neutron emission from deuterium, but the E_3 transition from $2 \cdot 50$ MeV to the ground state is above it. Any neutrons from this reaction, therefore, would be detected by a Szilard-Chalmers detector, employing sodium permanganate solution, which was placed in close proximity to the cylinders of heavy water. It was shown in a separate experiment that the detector was completely insensitive to any γ-rays from cobalt. After the experiment the sensitivity of the detector was measured using the standard $RaBeF_4$ neutron source and from the data the ratio of the number of transitions between the $2 \cdot 50$ MeV state and the ground state to the number of transitions to the $1 \cdot 33$ MeV state was determined as less than 1×10^{-6}.

V. References

Barnes, C. A., Carver, J. H., Stafford, G. H., and Wilkinson, D. H. (1952).—*Phys. Rev.* **86** : 359.

Bretscher, E., Cook, G. B., Martin, G. R., and Wilkinson, D. H. (1949).—*Proc. Roy. Soc.* A **196** : 436.

Broda, E. (1948).—*J. Chim. Phys.* **45** : 193.

Dodson, R. H., Goldblatt, M., and Sullivan, J. H. (1944).—Manhattan Dist. Declass. Cmttee. Rep. No. 344.

Edge, R. D. (1956).—Ph.D. Thesis, University of Cambridge.

Grace, M. A., Beghian, L. E., Preston, G., and Halban, H. (1951).—*Phys. Rev.* **77** : 286 ; **82** : 969.

Hollander, J. M., Perlman, I., and Seaborg, G. T. (1953).—*Rev. Mod. Phys.* **25** : 469.

Hughes, D. J., and Harvey, J. A. (1955).—Neutron cross sections. Brookhaven Nat. Lab. Rep. No. 325.

Kaplan, N., and Yagoda, H. (1952).—*Rev. Sci. Instrum.* **23** : 155.

Littler, D. J. (1952).—*Proc. Phys. Soc.* A **65** : 203.

Myers, V., and Wattenberg, A. (1949).—*Phys. Rev.* **75** : 992.

Schlögl, F. (1948).—*Z. Naturf.* **3** : 229.

Segre, E. (1952).—*Phys. Rev.* **86** : 21.

Segre, E. (1953).—“ Experimental Nuclear Physics.” Vol. 2. (Wiley : New York.)

Stafford, G. H. (1951).—*Proc. Phys. Soc.* A **64** : 388.

Szilard, L., and Chalmers, T. A. (1934).—*Nature* **134** : 494.

Tait, J. H. (1954).—*Proc. Phys. Soc.* A **67** : 615.

Titterton, E. W., and Brinkley, T. A. (1951).—*Proc. Phys. Soc.* A **64** : 212.

Wallace, P. R., and LeCaine, J. (1947).—Elementary approximations in the theory of neutron diffusion. Nat. Res. Counc., Canada. Pub. No. 1480.

THE BEHAVIOUR OF A CHAPMAN LAYER IN THE NIGHT F_2 REGION OF THE IONOSPHERE, UNDER THE INFLUENCE OF GRAVITY, DIFFUSION, AND ATTACHMENT

By R. A. Duncan*

[*Manuscript received July* 11, 1956]

Summary

It is shown that, in the presence of diffusion, gravity, and attachment, a Chapman layer, no matter what its height, maintains its shape, decaying uniformly with an effective attachment coefficient equal to the true attachment coefficient at the height of the electron density maximum ; and that, at the same time, the layer drifts bodily towards an equilibrium height.

It is then shown that a uniform vertical tidal drift will alter the equilibrium height of a Chapman layer.

I. Introduction

A number of workers (Martyn 1954 ; Duncan 1956 ; Yonezawa 1956) have suggested that the height gradient of electron decay, and diffusion under gravity should combine to stabilize the height of the night-time F_2 region of the ionosphere. If the layer is raised the increased diffusion under gravity should bring it down again, if it is lowered the increased decay of the lower edge should tend to lift the height of the electron density maximum.

Martyn (1956) gave a quantitative basis to these ideas by showing that, under the action of diffusion, gravity, and attachment, a Chapman layer at a reduced height (Z_m) such that

$$\beta=g\,\frac{\sin^2\psi}{2H\nu} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

decays uniformly without change of shape or height.

Here H is the scale height,
g is the gravitational field strength,
ψ is the geomagnetic dip,
ν is the positive ion collisional frequency,
β is the attachment coefficient.

By a Chapman layer is meant a region in which the height distribution of electron density is described by the relation

$$N=N_m\exp\tfrac{1}{2}[1-(z-z_m)-\mathrm{e}^{-(z-z_m)}], \qquad \ldots\ldots\ldots (2)$$

---

* Radio Research Laboratories, C.S.I.R.O., Camden, N.S.W.

where N is the electron density,
N_m is the maximum electron density,
z is the reduced height,
z_m is the reduced height of the electron density maximum.

For convenience, we shall take the height such that equation (1) is satisfied as the datum level, $z=0$, and the positive ion collisional frequency and attachment coefficient at this level will be denoted by ν_0 and β_0 respectively.

In this paper the conclusions reached by Martyn will be extended.

It will first be shown that, in the presence of diffusion, gravity, and attachment, a Chapman layer, no matter what its height, maintains its shape, decaying uniformly with an effective attachment coefficient (β') equal to the true attachment coefficient at the height of the electron density maximum (z_m); and that at the same time, the layer drifts bodily towards the equilibrium height ($z=0$) with a velocity

$$v=-\frac{2g\sin^2\psi}{\nu_0}\sinh z_m. \quad \dots\dots\dots\dots\dots\dots \quad (3)$$

It will then be shown that, if an ionospheric region, in equilibrium with gravity, diffusion, and attachment, is subjected to a vertical tidal drift W, its height is perturbed by an amount

$$\Delta z_m=\text{arc sinh}\,\frac{W\nu_0}{2g\sin^2\psi}=\text{arc sinh}\,\frac{W}{4H\beta_0}. \quad \dots\dots\dots \quad (4)$$

II. Solution of the Equation of Continuity for a Night-time Chapman Layer in the Presence of Gravity, Diffusion, and Attachment

We shall assume the attachment coefficient to be proportional to the pressure, that is,

$$\beta=\beta_0 e^{-z}. \quad \dots\dots\dots\dots\dots\dots\dots\dots \quad (5)$$

The continuity equation then becomes

$$\frac{\partial N}{\partial t}=\frac{2g\sin^2\psi}{H\nu_0}e^z\left(\frac{\partial^2 N}{\partial z^2}+\frac{3}{2}\,\frac{\partial N}{\partial z}+\frac{N}{2}\right)-\beta_0 N e^{-z}; \quad \dots\dots \quad (6)$$

Ferraro (1945); Martyn (1956).

Substituting equation (2) into this we get

$$\frac{\partial N}{N\partial t}=\frac{g\sin^2\psi}{2H\nu_0}e^z(e^{-2z}e^{2z_m}-e^{-z}e^{z_m})-\beta_0 e^{-z}. \quad \dots\dots\dots\dots \quad (7)$$

Now we have chosen our datum level such that

$$g\sin^2\psi/2H\nu_0=\beta_0, \quad \dots\dots\dots\dots\dots\dots\dots \quad (8)$$

and (7) becomes therefore

$$\frac{\partial N}{N\partial t}=\beta_0(e^{-z}e^{2z_m}-e^{z_m}-e^{-z}). \quad \dots\dots\dots \quad (9)$$

This may be expanded to

$$\frac{\partial N}{N\partial t}=2\beta_0(e^{z_m}-e^{-z_m})\tfrac{1}{2}\{e^{-(z-z_m)}-1\}-\beta_0 e^{-z_m}, \quad \ldots\ldots \quad \mathbf{(10)}$$

or from (**8**) and (**2**)

$$\frac{\partial N}{\partial t}=\frac{g\sin^2\psi}{H\nu_0}(e^{z_m}-e^{-z_m})\frac{\partial N}{\partial z}-\beta_0 e^{-z_m}N. \quad \ldots\ldots\ldots\ldots \quad \mathbf{(11)}$$

This equation corresponds to the general form

$$\frac{\partial N}{\partial t}=-v\frac{\partial N}{H\partial z}-\beta' N, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(12)}$$

where both v and β' are independent of the height z but depend only on the height of the maximum electron density of the layer z_m.

It is evident that the layer decays uniformly with an effective attachment coefficient

$$\beta'=\beta_0 e^{-z_m}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(13)}$$

and that the electron density maximum moves with a velocity

$$v=-\frac{g\sin^2\psi}{\nu_0}(e^{z_m}-e^{-z_m}) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(14)}$$

towards the height $z_m=0$, i.e. from the way in which we have defined our datum level, towards the height such that $\beta=g\sin^2\psi/2H\nu$.

III. The Behaviour of an Arbitrary Electron Density Distribution under Gravity, Diffusion, and Attachment

Martyn has suggested that, under the influence of gravity, diffusion, and attachment, an ionospheric region with any initial height distribution of electron density, approaches the Chapman form centred at the equilibrium height $z=0$.

Proof of this is still lacking, but we can show that all those height distributions of electron density that can be built up as the sum of a number of Chapman layers of different peak electron densities, and centred at different heights, approach the simple Chapman form. As equation (**6**) is linear, each of the component layers in such a region can be considered independently of the others and each therefore satisfies equation (**12**). Each of the component Chapman layers then, approaches the equilibrium height $z=0$, so that the electron distribution as a whole must approach the Chapman form, centred at the height $z=0$.

IV. The Perturbation of a Chapman Region by a Vertical Drift

If a Chapman region in equilibrium with gravity, diffusion, and attachment is subjected to a vertical tidal drift W it will move with the drift until the restoring velocity as given by equation (**14**) is equal and opposite to the perturbing tidal drift. It follows from these considerations that the perturbation of the equilibrium height of the layer is

$$\Delta z_m = \text{arc sinh}\,\frac{W\nu_0}{2g\sin^2\psi} = \text{arc sinh}\,\frac{W}{4H\beta_0}, \quad \ldots\ldots \quad (15)$$

where ν_0 is still the positive ion collisional frequency and β_0 the attachment coefficient, at the equilibrium height of the electron density maximum of the unperturbed layer.

V. Conclusions

Martyn's suggestion that the observed adoption of an approximately parabolic form by the night F_2 region is due to the action of gravity, diffusion, and attachment has been put on a firmer basis, although a proof that *any* initial electron distribution will adopt the Chapman form is still lacking.

Equation (**15**) should make possible a more concise formulation of the author's lunar tidal theory (Duncan 1956).

VI. Acknowledgments

The author is indebted to Dr. D. F. Martyn, the Chief Scientific Officer of the Radio Research Board, and to Mr. W. L. Price for helpful discussions.

VII. References

Ferraro, V. C. A. (1945).—*Terr. Magn. Atmos. Elect.* **50**: 215.
Duncan, R. A. (1956).—*Aust. J. Phys.* **9**: 112.
Martyn, D. F. (1954).—" The Physics of the Ionosphere." p. 256. (Phys. Soc.: London.)
Martyn, D. F. (1956).—*Aust. J. Phys.* **9**: 161.
Yonezawa, T. (1955).—*J. Radio Res. Labs. Japan* **3**: 1.

MOTION IN THE NIGHT-TIME E_s REGION AT BRISBANE

By J. A. Thomas* and M. J. Burke*

[*Manuscript received August* 6, 1956]

Summary

An analysis is made of night-time motion in the E_s region, as determined from a study of radio echoes at 2·28 Mc/s. Night-time E_s usually consists of moving clouds of ionization of lateral extent $\sim$10 km ; these clouds may be either isolated or close-packed to form layers of ionization. There is evidence that the clouds are sometimes elongated in a direction different from their direction of travel.

Good correlation is found between velocities of E_s movement as determined by fading analysis (Mitra 1949) and direction-finding techniques. Speeds of movement are grouped about 70 m/sec, and the winds are predominantly towards the north.

I. GENERAL

It has been shown in previous papers (Thomas and Svenson 1955 ; Thomas 1956*a*) that the two types of E_s at Brisbane (" sequential " type (E_{ss}), and " constant-height " type (E_{sc})), show considerable vertical motion of a tidal nature. It is logical to expect large horizontal motions to accompany the vertical movements in the ionosphere. Measurements of horizontal drift velocities of " patches " of E_s ionization have been reported from many parts of the world. In Australia, Harvey (1955) has measured the drift velocities of large patches of (mainly) sequential E_s as it occurs near Sydney.

An examination of the $P'f$ records at Brisbane shows that there are frequently E_{sc} traces which, over a period of time, gradually decrease and/or increase their range. Examples are shown in Figures 1 and 2. These changes are quite distinct from the large decrease of range associated with E_{ss} formation. Since the changes of range are the same at all frequencies up to fE_s, measurements made at any one frequency will be characteristic of measurements at all other frequencies ; use is therefore made of the various fixed frequency records which have been made at Brisbane by the techniques described by Thomas and McNicol (1955) and by McNicol, Webster, and Bowman (1956).

The various types of records available are outlined below :

(*a*) Simple $P't$ type records covering the ranges 0–500, 90–150, 200–500, 300–1000, 0–2800, 0–4000 km.
(*b*) Pulse-operated A.G.C. records of the $P't$ type in the range 90–150 km.
(*c*) Swept-gain $P't$ records, giving a measure of the relative intensity of echoes, and operating over the range 0–500 km.
(*d*) Phase-path records over the range 0–500 km.
(*e*) Records giving the directions of arrival of echoes in the range 0–500 km.

---

* Physics Department, University of Queensland, Brisbane.

(*f*) Fading records from a three-point aerial system (after Mitra 1949) giving values of ionospheric wind velocities.

(*g*) Simultaneous *P't* records (0–500 km) taken at stations about 100 km apart.

(*h*) Slightly oblique *P't* records (0–500 km) between stations about 100 km apart.

A double advantage arises from the use of such records in conjunction with the *P'f* records : the sensitivity and resolution of echoes is higher than with *P'f* records and the records are continuous in time, thus simplifying the interpretation of the various phenomena.

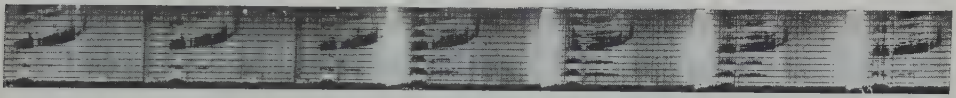


Fig. 1.—An E_s " cloud " on *P'f* records, showing successive decrease and increase in ranges.

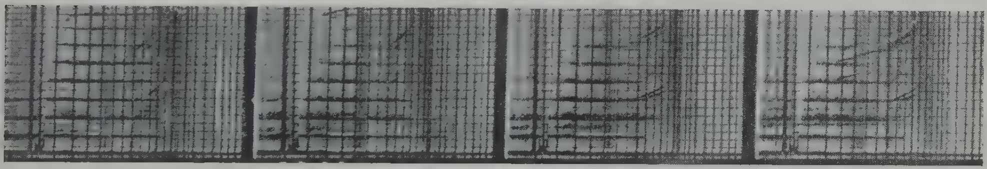


Fig. 2.—*P'f* records showing the gradual separation of a subsidiary E_s trace from the lowermost trace which remains at constant range.

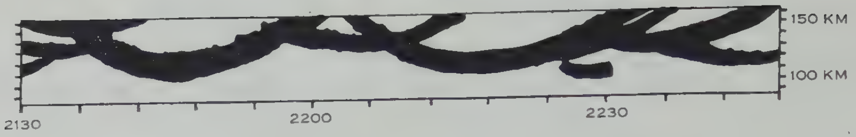
The greater part of the work described hereunder forms an analysis of night-time records obtained at 2·28 Mc/s at Brisbane. Some small use has been made also of *P't* records taken at 3·84 and 5·80 Mc/s.

II. NATURE OF THE PHENOMENA INVESTIGATED

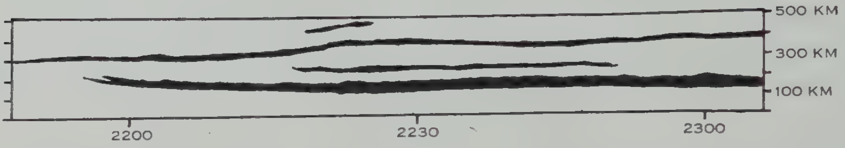
At 2·28 Mc/s the majority of E_s records are due to E_{sc}, although in the early parts of summer night-time records both E_{sc} and E_{ss} may occur either simultaneously or individually. At such times special " double-layer " phenomena are sometimes observed and these are discussed in some detail by Thomas (1956*b*).

Several slightly different types of *P't* records can occur, depending on whether a continuous background of E_{sc} is present or not. Tracings of the most common occurrences are shown in Figure 3. For convenience of classification these various types of record have been given the specific names mentioned in the caption to the figure. The distinction between a " layer " and a " cloud " is obviously only an artificial one ; a saucer-shaped trace lasting less than an hour is regarded as being due to reflection from the edge of a small isolated moving cloud rather than from a large sheet or layer of ionization, and, for convenience, the traces are given the same names as the assumed causes. Detailed examination of the expanded *E* region *P't* records (i.e. 80–160 km) shows that the " layer " type trace is quite often made up of a large number of

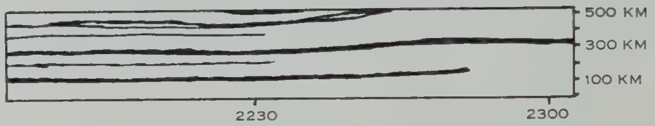
(a)



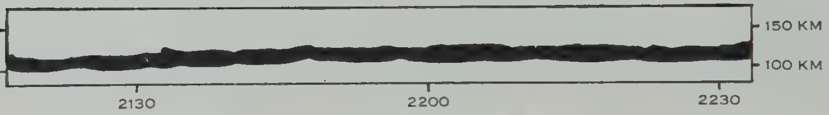
(b)



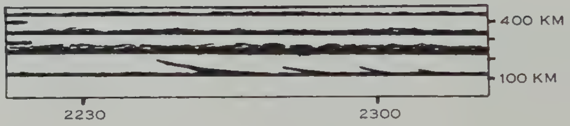
(c)



(d)



(e)



(f)

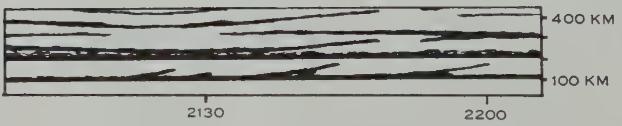


Fig. 3 (*a*).—Isolated E_s " clouds " recorded on expanded time-base $P't$ equipment. (Traces of this type are recorded as " clouds " when they do not exist for periods greater than one hour.)

Fig. 3 (*b*).—$P't$ record showing echoes from a " layer " of ionization moving " in " (or overhead). Such traces are recorded as " layer-in ", when the range remains sensibly constant after the initial decrease.

Fig. 3 (*c*).—$P't$ record showing echoes from a " layer " of ionization moving " out " (from overhead).

Fig. 3 (*d*).—Expanded $P't$ record showing the detailed structure of a " layer " of E_s ionization.

Fig. 3 (*e*).—$P't$ record showing traces converging to the main E_s trace. Such traces were recorded as " spurs in ".

Fig. 3 (*f*).—$P't$ record showing traces diverging from the main E_s trace. Such traces were recorded as " spurs out ".

overlapping " cloud " type traces (Fig. 3 (*d*)). Sometimes the more extreme ranges of these clouds are masked, but at other times traces may be seen entering and leaving the base trace giving rise to the " spurs " of Figures 3 (*e*) and 3 (*f*).

The critical frequency of such E_s spurs, which show as apparent stratified layers on $P'f$, may be slightly different from that of the base layer (either greater or less) but the difference is usually less than 2 Mc/s. This lends support to the view that the E_{sc} region consists of clouds which may or may not be close enough to give continuous echoes on $P't$; one would not expect all such clouds to have exactly the same ionization density.

The phenomenon of frequency spreading in F region echoes has been discussed by Singleton (1956) and he arrives at a similar conclusion as to the cause. In the E_s region, such frequency spreading can only be observed when it is accompanied by range stratification.

Very occasionally, traces are recorded in which no discernible evidence is shown of a range decrease at the beginning or a range increase at the end of the trace. The rate of fading of amplitude in such traces is always found to be very slow, indicating very little ionospheric movement; there is no preferential time for the appearance of such traces.

III. Nocturnal and Seasonal Distribution of the Phenomena

The nocturnal distributions of the times of occurrence (i.e. start and finish) of clouds, spurs, and layers have been separately and collectively plotted for each month of 1952. As would be expected from the argument of the previous section, there is no major disagreement between the three phenomena. Separate clouds are more easily identifiable in the base layer on some occasions than on others, so that more clouds are recorded than spurs; at other times the reverse is true. The nocturnal distributions for all the phenomena are usually similar in any one month.

The collective histograms show a tendency for more apparent E_s movement to occur in the summer months before midnight than after. It should be remembered that sequential E_s persists until around 2100 hr in summer; this may well be the reason for the enhanced recordings of apparent movement.

It has been found (Thomas 1956*a*) that the occurrence of night-time E_s echoes at 2·28 Mc/s is fairly constant throughout the year. The histograms show, however, that in October 1952 there was little apparent evidence of E_s motion—E_s was present for about 60 per cent. of the night but sloping traces were not observed as frequently as during the other months. No real explanation can be put forward for this fact, but it has been found (Section V) that during this particular month the " winds " in the E_s region at night tended to maintain a roughly constant amplitude with a smooth change in direction; this may possibly be a significant factor in explaining the paucity of sloping $P't$ traces.

IV. Methods of Measurement of Speeds and Velocities

Since there is very little correlation between $P't$ records taken at points 50 km apart, all calculations of speeds and directions of E_s movements have been made from data recorded at one station only, using the curved traces referred to in the previous sections.

In making most of these calculations one is forced to put forward some simple hypothesis regarding the E_s patch and its motion. The simplest hypothesis is to suppose that the E_s patch may be regarded as a point reflector moving with uniform horizontal velocity. The effect of changes in the nature of the patch will be shown in Section VI.

(*a*) *Speeds*

There are three basic methods for determining speeds from single-station records. Consider the diagrams of Figure 4. While these diagrams are not necessarily in a vertical plane through the transmitter, it is here assumed, following the above hypothesis, that v is purely horizontal.

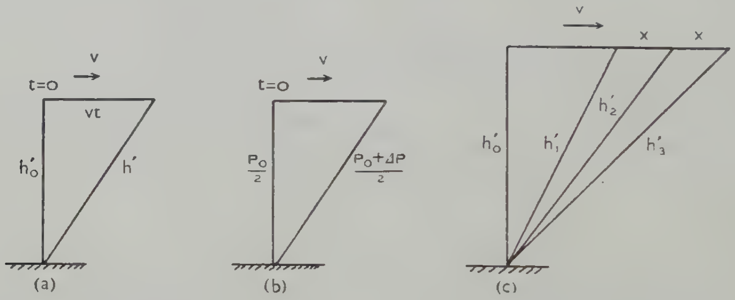


Fig. 4.—Diagrams illustrating three methods of calculating the horizontal speed of cloud movement.

(i) We have (Fig. 4 (*a*))

$$v^2=\frac{h'}{t}\cdot\frac{\mathrm{d}h'}{\mathrm{d}t}, \qquad \ldots\ldots\ldots\ldots \qquad (1)$$

or

$$v=\frac{h'}{\sqrt{(h'^2-h_0'^2)}}\cdot\frac{\mathrm{d}h'}{\mathrm{d}t}. \qquad \ldots\ldots\ldots\ldots \qquad (2)$$

Either of these expressions may be used to derive the speed of movement of the patch. The latter is to be slightly preferred, as the minimum range h_0' can often be determined more easily than the exact time at which minimum range is reached (i.e. the instant $t=0$).

(ii) Here (Fig. 4 (*b*)) we have (after Findlay 1953)

$$\Delta P=2v^2t^2/P_0.$$

A plot of ΔP as a function of t^2 thus gives a measure of v, if we assume that for night-time records very little error is incurred by substituting $P'_{\text{min.}}$ for P_0 in the above expression.

(iii) In this case (Fig. 4 (*c*))

$$x^2=\tfrac{1}{2}(h_3'^2+h_1'^2-2h_2'^2)$$

so that by measuring h' at definite intervals one may find v. This method is only suitable where very accurate measurements of h' are available (i.e. to better than 0·5 km)—large errors arise in the squaring and further manipulation if the readings are not very accurate.

(*b*) *Velocities*

The additional directional data necessary to determine the velocity of movement of E_s patches is provided by the automatic recording of the direction of arrival of the echoes. Knowing the azimuth and elevation of the reflecting point and its slant range, the point of reflection may be plotted at 3-min intervals on a horizontal polar diagram (at the E_s level) and the horizontal motion of the reflecting point easily calculated.

It will be shown in Section VI that for E_s clouds as observed at 2·28 Mc/s at Brisbane accurate velocity measurements can be obtained from the directional data even in the case of clouds which have elliptical sections in horizontal planes but which are assumed to be negligible in vertical thickness. The correct motion is still found even when the elliptical cloud is orientated to this direction of motion ; it is only necessary to neglect (for this purpose) any readings giving the reflecting point less than 10 km from the zenithal point at the E_s level and to thereby isolate the decreasing and increasing range portions of a cloud. This sacrifice is not very great, as it usually only means the non-utilization of one or two D.F. readings.

It is not possible to obtain directional data on all E_s clouds as the signal strength from the oblique echoes is often quite low, and at 2·28 Mc/s E_s nearly always shows considerable amplitude and phase fading. These two facts often make an accurate measurement of the direction of an echo impossible. Nevertheless, some 60 clouds or spurs have been so examined from about 15 weeks' total running time spread throughout the year.

In addition to this method of finding the drift velocity of E_s patches, there is the method of fading analysis of three-point amplitude records developed by Mitra. This method gives the velocity of drift of the ionospheric diffracting screen, but some doubt often exists as to the position of the diffracting screen in the ionosphere. Special measurements taken at Brisbane to check this point by correlating the direction-finding velocities with the Mitra " winds ", have shown that for the night-time E_s region the correlation is good. The results of this series of measurements are shown in Table 1. Such disagreement as exists can be quite easily explained if one remembers the simplifying assumptions of the Mitra technique, namely, that " the line of maximum amplitude " is normal to the drift direction and that the drift velocity is large compared with the velocity due to random changes. In the former case, if the normal to the line of maximum amplitude is inclined at an angle θ to the direction of the drift velocity, the resultant velocity will be too low in magnitude by a factor $\cos\theta$ and in error in direction by the value θ. In the latter case Yerg (1955) has shown that, if the velocity due to random changes is of the same order as the drift velocity, then the " method of similar fades " as used in Brisbane will give results much greater in magnitude than the true drift velocity (by up to several hundred per cent.), and approximately correct in direction. It is thus highly likely that there will be some errors in the values obtained from each E_s patch used in the above correlation tests.

In view of the correlation between these two methods of drift determination and the paucity of good direction-finding records, nearly all the velocity results are taken from the analysis of the fading patterns.

Table 1

CORRELATION BETWEEN VELOCITY MEASUREMENTS

Date	Time	Direction of Arrival Velocity		Mitra Velocity		Errors	
						Speeds (%)	Angle
		(m/sec)		(m/sec)			
5.iv.56	1912	75	315°	64	350°	—15	+35°
	1939	45	20°	31	61°	—30	+41°
	2003	73	210°	(200)	(158°)	+180	—52°
	2021	100	15°	60	31°	—40	+16°
	2100	90	310°	104	355°	+13	+45°
	2239	(75)	(345°)	86	327°	+15	—18°
	2355	110	274°	171	244°	+55	—30°
9.iv.56	2300	(60)	(98°)	56	9°	—7	—89°
10.iv.56	2230	50	304°	58	332°	+16	+26°
11.iv.56	2055	100	320°	100	336°	0	+16°
11.iv.56	2020	100	5°	100	(17°)	0	+12°
13.iv.56	0400	80	20°	67	325°	—16	—55°
16.iv.56	2030	105	300°	111	334°	+6	+34°
21.iv.56	0340	54	305°	60	319°	+10	+14°
				Average	.. ..	+13	+1·5°

V. Results

(*a*) *Speeds and Cloud Sizes*

A histogram of speeds derived by the methods of Section IV is shown in Figure 5. Although there is a large spread in values, the most probable speed appears to lie in the vicinity of 4·5 km/min and the mean speed is about 5 km/min.

Speed measurements may be coupled with observations of blanketing by clouds (Fig. 6) to give an estimate of cloud size. The period of zero phase-path change may be correlated with cloud speeds in the same fashion. Thus, for example, the estimated speed of the cloud of Figure 6 is 3·9 km/min and blanketing occurs for 2·5 min ; this leads to a cloud about 10 km in extent in the direction of movement. The average cloud size as determined by 28 such measurements is 13 km. This value agrees well with the lack of correlation between records taken 50 km apart, as reported in the earlier paper.

(*b*) *Velocities*

Although fading records have been taken on the *E* region for both day and night-time, only the night-time results are included herein, as at the operating frequency of 2·28 Mc/s it is often impossible to determine whether the day-time reflection is from the E_s or the normal E region. All wind directions quoted are vector directions.

In Figure 7 are shown polar plots of the hourly changes in E_s drift velocity for the year July 1952 to June 1953. For nearly all months the winds have an easterly component in the early hours of the night followed by a swing to the west. Harmonic analysis* shows that this trend is due to the semi-diurnal component of the east-west variation ; for nearly all months this has a maximum amplitude in the easterly direction at about 18–20 hr.

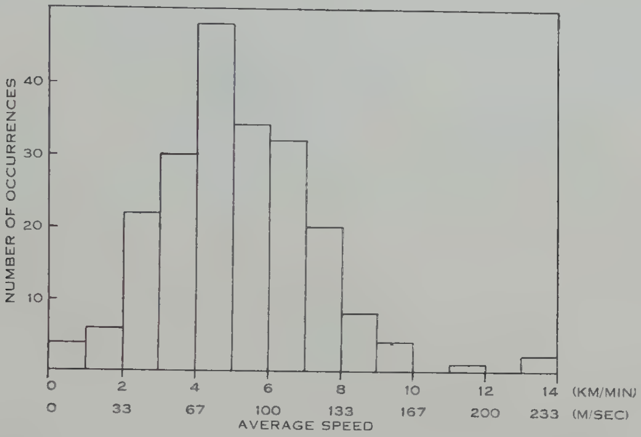


Fig. 5.—Distribution of E_s speeds.

Figure 8 shows the prevailing winds for night-time E_s for this same period. It is to be noted that the majority of these prevailing winds tend to be northerly with a bias towards the west.

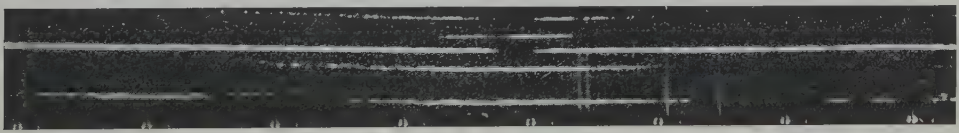


Fig. 6.—An example of F-blanketing caused by an E_s cloud. Measurements of the cloud speed and duration of blanketing give an estimated cloud size of 10 km.

The velocities as found from the directional data are plotted as a scatter diagram in Figure 9. As mentioned previously, these data represent readings taken over a long period and there has been no attempt to separate the small number of values into diurnal or seasonal components. One obvious and striking fact, however, is that there were very few velocities recorded with a southerly component ; there is a preponderance of velocities in the north-west sector. This diagram is thus in good agreement with the prevailing winds derived from fading analysis and shown in Figure 8.

---

* The harmonic analysis is carried out using data from the full 24 hr ; this work will be reported in a further paper.

VI. Cloud Shapes

Measurements of the vertical thickness of E_s layers may be made by comparing the measured height of M-echoes with the height calculated on the basis $2h'F-h'E_s$. Such measurements have been made at Brisbane from the $2\cdot28$ Mc/s $P't$ records using expanded range scales (Hosking, unpublished data). The results show that night-time E_s has a thickness which is probably not greater than 1 km.

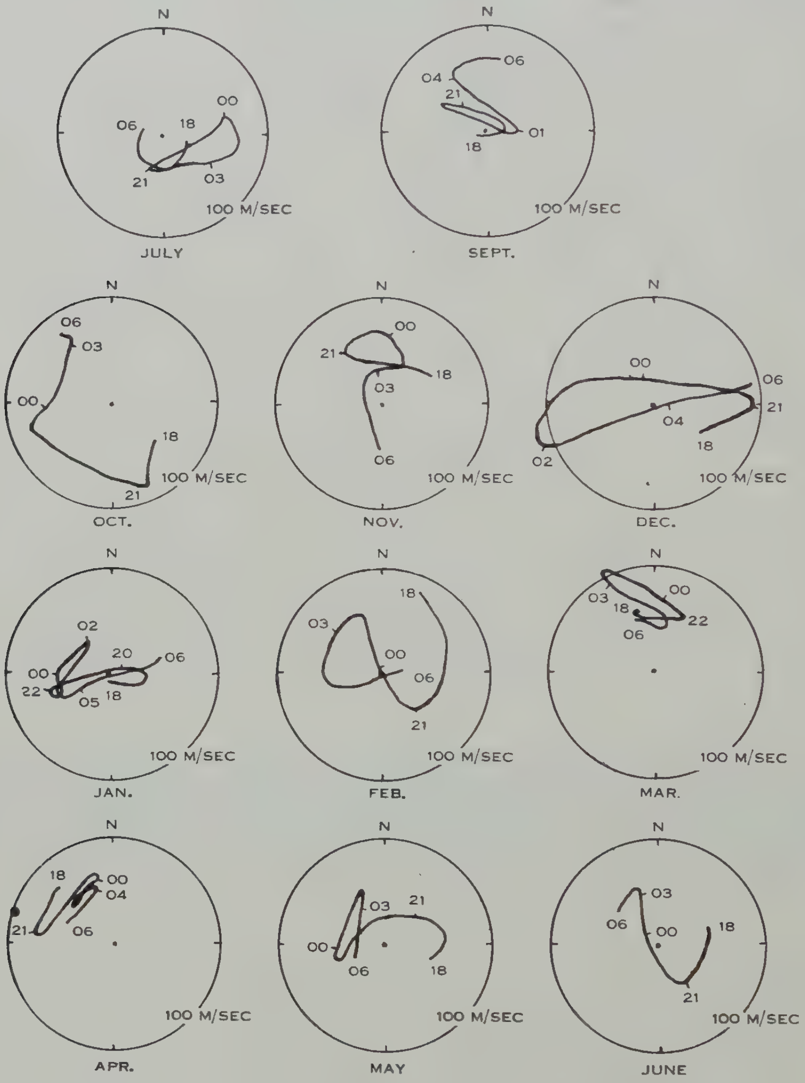


Fig. 7.—Diurnal variation of velocity of drift of night-time E_s for the year July 1952–June 1953.

The measurements of Section V (a) indicate that the probable horizontal extent of an E_s cloud is of the order of 10 km.

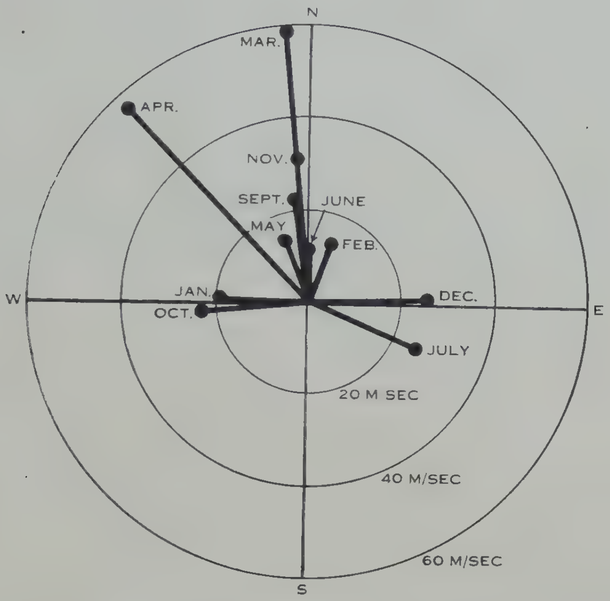


Fig. 8.—Prevailing wind components of night-time E_s drift velocities for the year July 1952–June 1953.

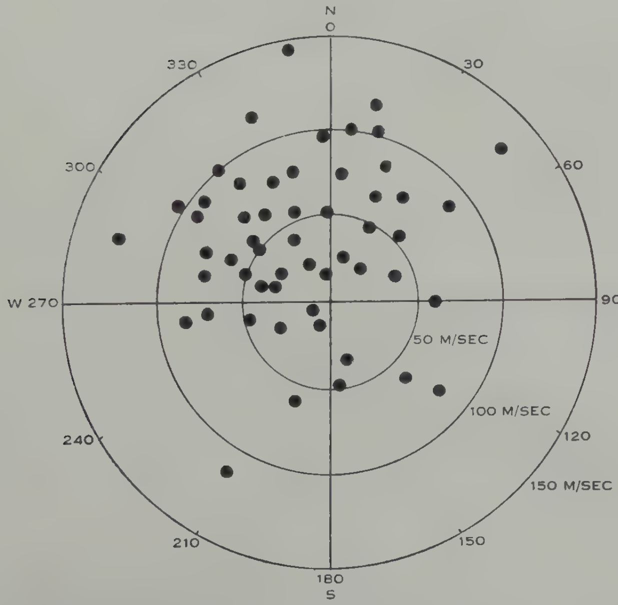


Fig. 9.—Scatter plot of E_s cloud velocities as determined from directional recordings.

It is necessary to examine what effect, if any, the shape of a cloud will have on the velocity calculated by the methods outlined above.

(*a*) If the cloud is spherical of diameter 1 km the velocity values will be essentially correct, since it would only be necessary to add $\frac{1}{2}$ km to all range values to have the motion of the " point " cloud of the original hypothesis.

(*b*) If the cloud is cylindrical (10 km diameter, 1 km high) then it is easy to calculate what the resultant effect will be on the range and directional data. If we choose a cloud speed of 5 km/min (83 m/s) and a variety of paths varying from those passing vertically overhead to those at some degree of obliquity, the relevant data are shown plotted on the diagram of Figure 10. If the edges are rounded off to $0 \cdot 5$ km curvature, the diagram is not noticeably altered.

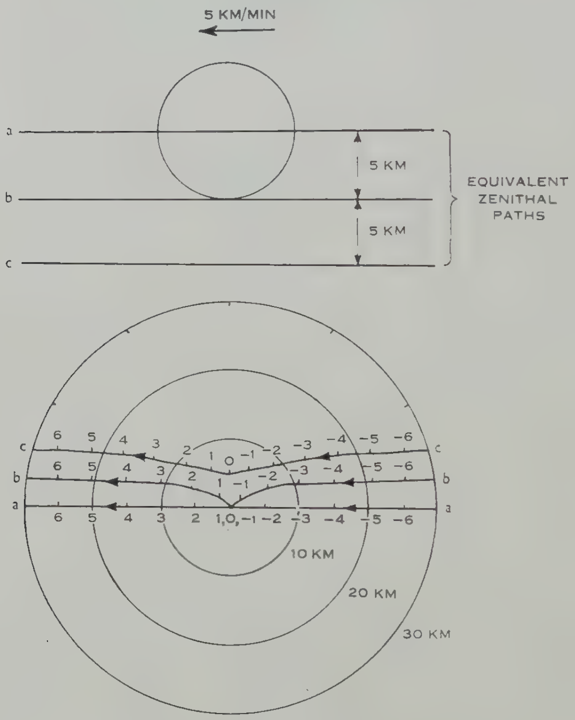


Fig. 10.—Reflection point paths for a circular disk cloud travelling at 5 km/min. The numbers on each path represent time in minutes ; the time $t=0$ is taken as the time of closest approach of the cloud centre to the zenith.

(*c*) If the cloud is elongated in one direction, then besides the variation in the ratio of diameters we have to consider also the possibility of movement of the cloud at an angle to the direction of elongation. To put this on a more concrete basis, we may once again choose a speed of

5 km/min and plot the positions of the reflection points for elliptical cylinders of varying magnitude, eccentricity, and orientation. The plot for one case in which the elliptical cylinder moves in a direction different from that of its major axis is shown in Figure 11.

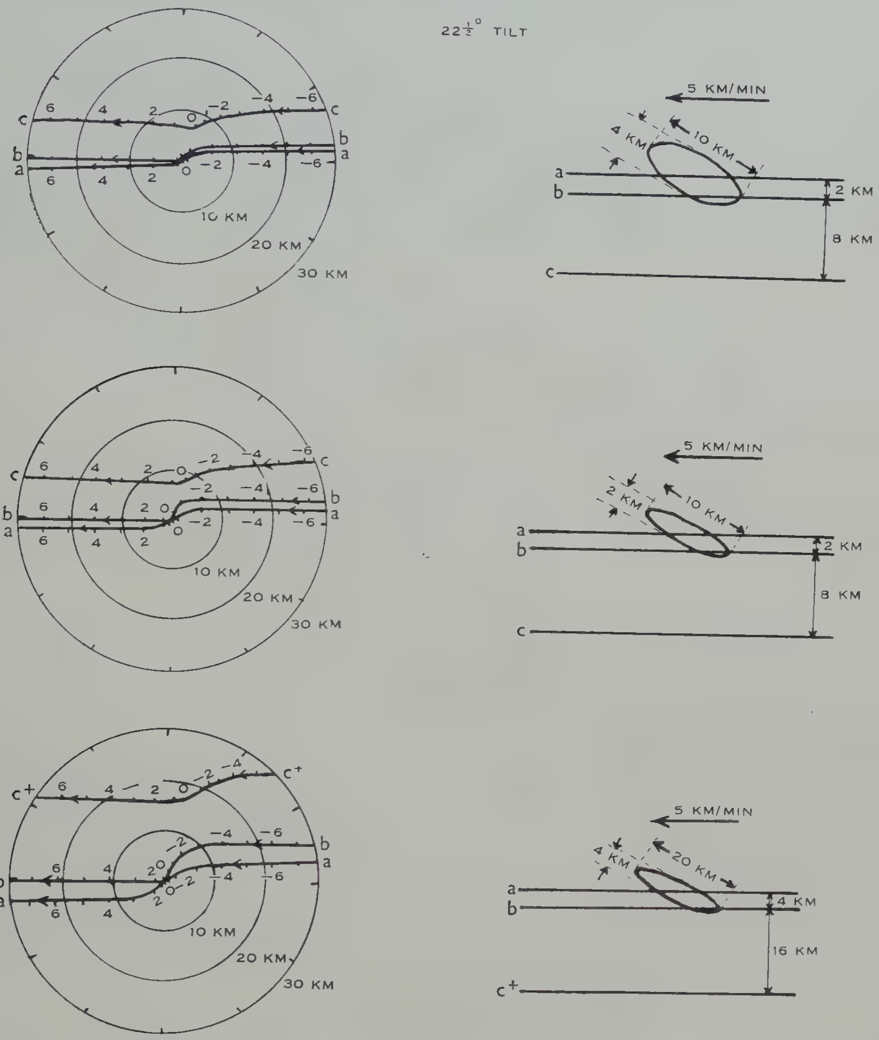


Fig. 11.—Plots of paths of reflecting points for reflection from the edge of elliptical cylinders for the configurations, paths, and speeds, shown on the right-hand side. *a*, *b*, and *c* represent different effective paths of the zenithal point past the cloud. Time is marked on the paths in minutes, zero time being taken as the time when the cloud centre is nearest the zenith.

From these diagrams it is clear that if we are to find the true drift velocity, we must deal separately with the decreasing and increasing range portions of the data, and avoid using the data which indicate that the reflecting point lies within a circle of radius ~10 km about the zenithal point at the cloud level.

If these precautions are taken the residual errors are negligible. If accurate directional measurements could be made within the 10 km zone, information could be derived concerning the eccentricity of the cloud. Such measurements would be extremely difficult to achieve in practice, owing to the fact that the cloud echoes are weak and often subject to severe phase and amplitude fading which limits the accuracy with which the direction of arrival may be determined. Some information may, however, be obtained from a close examination of the expanded *P′t* records. If we suppose that the E_s cloud of Figure 11 (c^*) is at a height of 100 km and calculate the slant range to the reflecting point on the cloud at various times, we obtain the hypothetical *P′t* curve of Figure 12. The striking point is the lack of symmetry of the " trace " about the point of minimum range (which does not occur at our time $t=0$). A calculation of speed from this curve shows that nearly correct speeds are found only from the steeper side of the trace ; speeds calculated using the other side of the trace are all too small ($\sim 3 \cdot 5$ km/min).

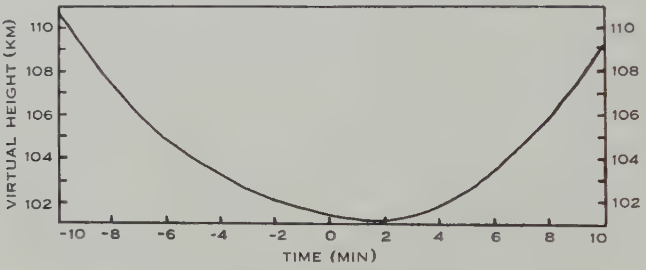


Fig. 12.—*h′t* plot of cloud (c^*) of Figure 11.

Such asymmetric traces do occur in the *P′t* records, although the asymmetry is often very slight ; in such cases the higher speed has been used for statistical purposes. Some of the cloud traces of Figure 3 clearly show asymmetry.

The conclusion is forced, then, that E_s clouds can and often do have considerable elongation, and that the elongation may be in a direction different from the direction of travel of the cloud. The elongation is subject to the overall restriction that the mean cloud size is ~ 10 km.

VII. CONCLUSIONS

It has been demonstrated that E_{sc} consists of (generally) moving patches of ionization ; the average speed of movement is about 80 m/sec, and the directions of drift have semi-diurnal swings from the north-east quadrant in the early evening to the north-west quadrant after about 22–2300 hr. Prevailing winds tend to be towards the north.

Layers of E_{sc} ionization are thought to be aggregates of roughly horizontal disk-like clouds, the disks not necessarily being circular, but at times showing considerable eccentricity in a direction differing from the direction of travel.

VIII. ACKNOWLEDGMENTS

This work forms part of the programme of the Radio Research Board of C.S.I.R.O. and is published by permission of the Board. The authors wish to thank Professor H. C. Webster, Mr. R. W. E. McNicol, and Mr. G. G. Bowman for their advice and discussion throughout the work.

IX. REFERENCES

FINDLAY, J. W. (1953).—*J. Atmos. Terr. Phys.* **3**: 73.
HARVEY, J. A. (1955).—*Aust. J. Phys.* **8**: 523.
McNICOL, R. W. E., WEBSTER, H. C., and BOWMAN, G. G. (1956).—*Aust. J. Phys.* **9**: 247.
MITRA, S. N. (1949).—*Proc. Instn. Elect. Engrs.* III **96**: 441.
SINGLETON, D. G. (1956).—M.Sc. Thesis, University of Queensland.
THOMAS, J. A. (1956*a*).—*Aust. J. Phys.* **9**: 228.
THOMAS, J. A. (1956*b*).—*Aust. J. Phys.* **9**: 574.
THOMAS, J. A., and McNICOL, R. W. E. (1955).—*Proc. Instn. Elect. Engrs.* **102B**: 793.
THOMAS, J. A., and SVENSON, A. C. (1955).—*Aust. J. Phys.* **8**: 554.
YERG, D. G. (1955).—*J. Geophys. Res.* **60**: 173.

HIGH-FREQUENCY BACKSCATTER OBSERVATIONS AT SALISBURY, SOUTH AUSTRALIA

By C. G. McCue*

[*Manuscript received March* 15, 1956]

Summary

An improved pulse technique which permits simultaneous observation of time delay and elevation angle of arrival of ionospheric echoes has enabled recognition of seven different modes of backscatter propagation involving both ionospheric and ground origins. By the use of four transmitting aerials with differing directional characteristics, it is shown that land is a more prominent source of backscatter than sea at low angles of elevation and that extremely large changes in land roughness can be identified by the resultant increase of backscatter echo amplitude.

I. Introduction and Historical Survey

If, for frequencies greater than the critical penetration frequency of the most densely ionized ionospheric layer, radio wave propagation via the ionosphere were limited to a series of specular reflections at the ionosphere and at the Earth's surface, no radio wave energy, other than that of the ground wave, would be detectable within the zone around the transmitter known as the skip zone. Nevertheless, it often has been possible to detect signals within the skip zone. In particular, signals have been detected by a receiver placed very close to the transmitter and employing the same aerial system. This apparently non-specular signal component has been named backscatter.

Theoretically, backscatter could originate in many ways, namely,

(i) Some of the upgoing wave energy may be scattered to points within the skip zone by irregularities in the E region of the ionosphere, e.g. sporadic-E ionization or the ionized trails of meteors.

(ii) After reflection at the level of the E region, the downgoing wave may be partially scattered by irregularities on the Earth's surface and part of this scattered radiation may be returned to a receiver in the skip zone via the E region.

(iii) Backscatter may be returned from irregularities on the Earth's surface after the wave has undergone more than one reflection at the E region, i.e. higher order or multiple-hop echoes of the type (ii).

(iv) Backscatter may arise as in (ii) with the F_2 layer being the reflecting medium.

(v) Multiple-hop backscatter of type (iv) may occur.

(vi) After reflection at the F_2 layer, part of the downgoing wave energy may be backscattered by irregularities in the E region.

---

* Electronics Research Division, Weapons Research Establishment, Salisbury, S. Aust.

(vii) Backscatter may be caused by the scattering of part of the upgoing wave energy by irregularities in the F_2 region.

(viii) Energy may be backscattered, at frequencies higher than the critical penetration frequency of the F_2 layer (f_0F_2), from irregularities above the height of the maximum ionization density of the regular F_2 layer.

(ix) Combinations of all the above modes may operate to produce backscatter, e.g. an upgoing wave may be scattered and deviated by irregularities in the E region and the scattered energy may travel up to the F_2 layer from whence it may be reflected back to ground within the skip zone.

These various backscatter types, (i)-(ix), are illustrated in Figure 1.

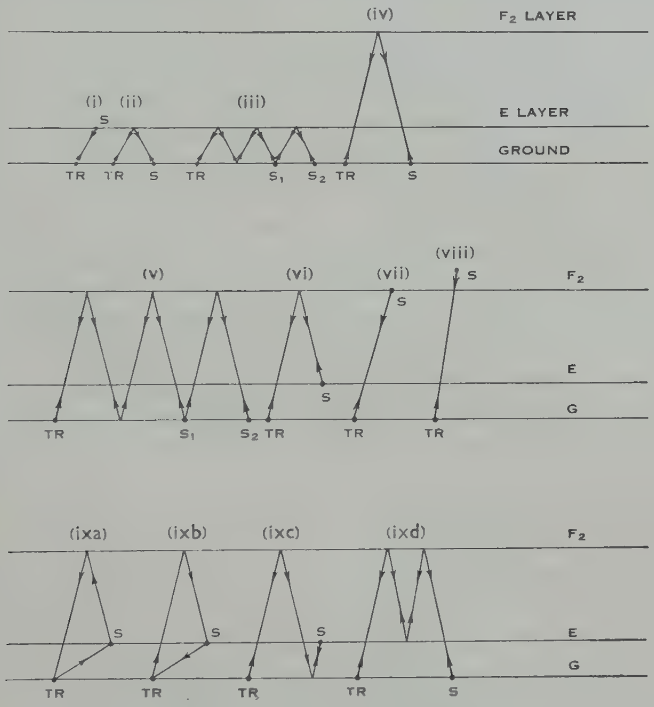


Fig. 1.—The possible sources of backscatter. *TR* denotes transmitter/receiver and *S* is the scatter source.

Backscatter was first observed about 1926, but it was almost neglected until Eckersley (1940) published a lengthy paper suggesting that backscatter could be divided into two classes, short and long scatter. The short scatter was supposed to arise in the manners (i) and (ix*b*) shown in Figure 1 and the long scatter in the manner (vi). Eckersley had neglected several factors in his work and Edwards and Jansky (1941) proposed, contrary to Eckersley, that the majority of long scatter arose in the manner (iv). However, the results of Edwards and Jansky were not regarded as conclusive owing to the inherent inaccuracy of their measurement technique.

Papers have been published subsequently reporting the occurrence of the various backscatter types. Most authors are agreed upon the origins of short scatter, stressing types (i) and (ii) as the most important. Regarding long scatter, Eckersley, Millington, and Cox (1944) rather inadequately supported the theory that type (vi) is the most prominent and Benner (1949), Peterson (1949, 1951), Hartsfield, Ostrow, and Silberstein (1950), Kono (1950), Abel and Edwards (1951), Dieminger (1951), Hartsfield and Silberstein (1952), and Silberstein (1953, 1954) supported the theory that type (iv) is the most prominent. The papers of Dieminger, Abel and Edwards, and Silberstein are most significant and will be discussed further.

Abel and Edwards (1951), operating on frequencies of 9, 12, 16, and 22 Mc/s, compared the echo from a transponder with long scatter. They used four transponders at distances of 700, 1000, 1500, and 2000 km from their transmitter/receiver. As the ground range of one transponder for one of the four frequencies passed through the skip distance, the transponder and backscatter echoes merged, indicating a common origin, the ground. This was the principal result of this paper. However, some cases of disagreement with the common ground origin were noted and Abel and Edwards attributed these to further causes.

Silberstein (1954) used a sweep-frequency technique which displayed the backscatter over a wide frequency range on the same record as the normal vertical incidence ionospheric sounding. This is illustrated in Figure 2 where the expected characteristics of types (iv) and (vi) backscatter are shown in relation to the normal vertical incidence data. Under usual ionospheric conditions, Silberstein principally obtained records conforming with the typical type (iv) behaviour. His results were difficult to interpret during ionospheric disturbances, but they suggested a number of cases where backscatter was apparently type (vi).

Dieminger (1951) has reviewed backscatter research up till 1951. He also described his own results obtained using two experimental methods. His principal technique was the sweep-frequency experiment described above. Using this, he compared the observed backscatter characteristics with those theoretically expected for the different types. He made additional observations over long periods of the behaviour of backscatter on fixed frequencies. He identified many of the backscatter types, but his results showed type (iv) to be the dominant long scatter type in winter at least.

All previous reliable backscatter data have been obtained in the northern hemisphere, principally during winter and between latitudes 40 and 60°, and on frequencies above the F_2 layer critical penetration frequency. These data suggest that the most common form of scatter observed in winter is type (iv). Few summer observations have been reported and these suggest that type (i) scatter is relatively common then, especially during the midday period.

Facilities existed at the Weapons Research Establishment which, with little difficulty, were capable of being adapted for studying backscatter. Equipment, including a high power transmitter in conjunction with four differing aerial systems and a cathode-ray direction finder, was made available on a part-time basis to the present author to enable him to examine several outstanding problems in connexion with backscatter. Principal among these problems were the identification and occurrence ratio of the various scatter types during summer and winter at a temperate latitude southern hemisphere station, a comparison of scatter from land and from sea, and the possible detection of scattered echoes from prominent land masses, such as mountains. Further, due to the available frequencies, it was possible to examine scatter below the F_2 layer critical penetration frequency.

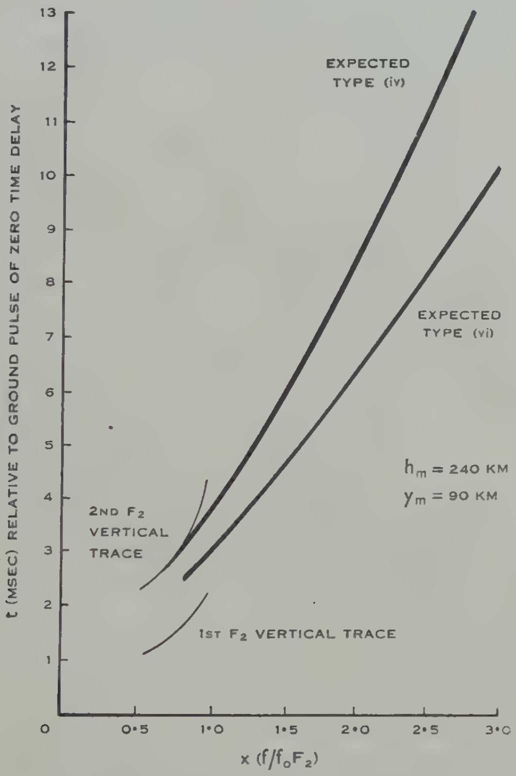


Fig. 2.—Theoretical behaviour of long scatter using sweep-frequency technique. (h_m is height of F_2 layer maximum electron density and y_m is the semi-thickness of the layer.)

II. Equipment and Experimental Programme

A programme of measurement of the time delay t relative to a ground pulse of zero delay and of the angle of elevation, α, of the downcoming backscatter was commenced at Salisbury in October 1954. Observations were made on 28 days during October–December 1954 (summer) and on 40 days during May–July 1955 (winter). Most observations were made during the hours 1200–1400 L.M.T., but some observations were made during all hours from 0930 to 1930 L.M.T. The summer programme was carried out using only one transmitting aerial, while, during winter, four different aerials were used.

The outstanding feature of the adopted technique for examining backscatter was the comparative ease with which the various scatter types could be recognized. Two distinct sets of circumstances were encountered, namely, when F_2 vertical returns were present and when they were absent. The frequency and observational conditions were so chosen that the former condition usually prevailed. This meant that the theoretical and observed properties of backscatter could be related very easily to the properties and time delays of the observed F_2 verticals independently of any subsidiary vertical incidence ionospheric soundings. Also, observational conditions were selected so that the periods when no F_2 verticals were recorded were immediately preceded by periods when F_2 verticals were present. Consequently, it was always possible to follow the different backscatter type echoes through the transition from one set of conditions to the other. These theoretical properties of the various backscatter types have been discussed at great length in the literature (Eckersley 1940; Dieminger 1951; Peterson 1951; McCue 1955) and they will not be described again here.

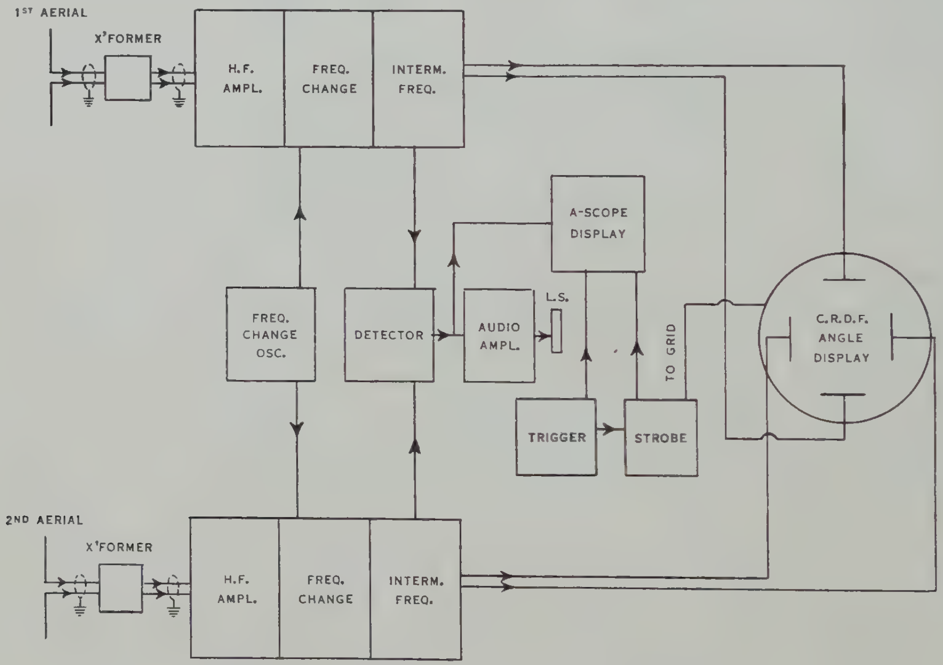


Fig. 3.—Block diagram of receiving equipment.

Attention was focused during the Salisbury experiments on scatter types (i)–(vii). It is possible that other scatter types occurred which escaped detection owing to attention being confined to the seven selected types which were expected to be most common.

The transmitter produced pulses the low frequency envelope of which was a half-sine wave of 300–600 μsec duration at a pulse repetition frequency of 50 c/s with a mean power of several hundred watts. The receiving system consisted of two vertically spaced parallel horizontal 30 ft dipoles at heights of 40 and 70 ft above the ground and coupled to the twin receiver channels of a cathode-ray direction finder (CR.DF) which is essentially a phase and amplitude comparator (de Walden *et al.* 1947). The CR.DF display was intensity-modulated by a strobe of variable width and time delay, enabling any specific part of the backscatter echo pattern to be examined separately. The strobe was added also to the time base of an A-scope display which was fed from a detector stage of the CR.DF—see Figure 3. The differing polar diagrams of the two dipoles enabled the elevation angle to be measured on the CR.DF, while t was recorded from the A-display unit. All observations were made visually. The CR.DF display was calibrated against actual angle of elevation using a signal transmitted from an aeroplane whose position was measured using graduated binoculars. The CR.DF angle calibration curve was found to be independent of azimuthal direction.

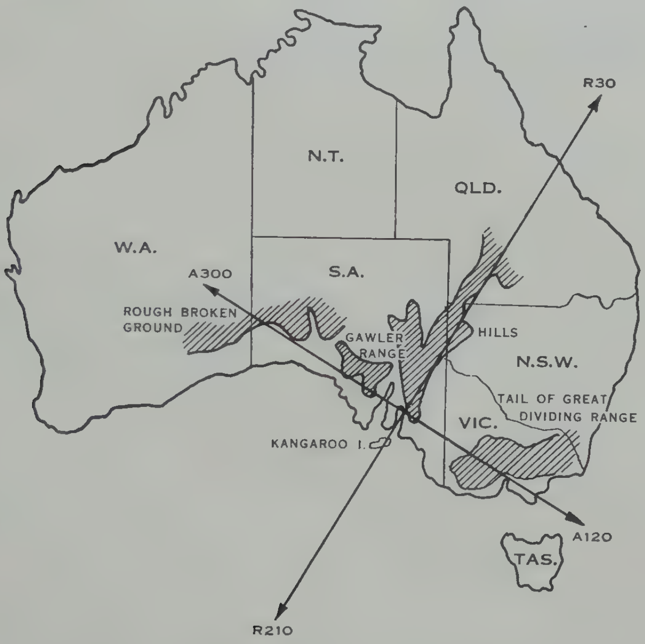


Fig. 4.—Map of Australia showing the orientations of the transmitting aerials used at Salisbury. The shaded areas indicate hilly or mountainous country.

The four transmitting aerials used during these experiments were:

(1) An array aerial which gave a unidirectional beam 15° wide in azimuth and about 70° wide in elevation. Its side lobes were down at least 10 dB on the main beam which was directed 120 °E. of N. from Salisbury, i.e. towards Melbourne. This was the aerial used during summer and will be called the A120 aerial throughout this paper.

(2) The above array reversed to point 300° E. of N. from Salisbury, i.e. towards the border of South Australia and Western Australia (denoted by A300).

(3) A rhombic aerial which was designed to have its maximum radiation directed at 23° to the Earth's surface and towards 210° E. of N. from Salisbury, i.e. over Kangaroo Island towards the Southern Ocean (R210).

(4) The above rhombic reversed to point 30° E. of N., i.e. towards Broken Hill (R30).

The orientations of these aerials are depicted in Figure 4, where the shaded areas on the map indicate hilly or mountainous country which may be prominent sources of backscatter. The rhombic polar diagram was not as simple as that of the array and allowance had to be made for this when interpreting the observations obtained using the rhombics.

III. THE OCCURRENCE OF THE VARIOUS SCATTER TYPES

Seven backscatter types, (i)–(vii), were observed at Salisbury during summer using the A120 transmitting aerial and during winter using the four aerials, A120, A300, R30, and R210. Type (iv) was by far the most commonly recorded scatter type during both seasons, independently of the aerial used and of the presence of F_2 verticals. Type (v) was the next most commonly observed type when F_2 verticals were present, but, when no F_2 verticals were to be seen, type (i) was more common than type (v). The rarest scatter type was (iii) followed by (vi), (vii), and (ii) in increasing rate of occurrence. Types (i) and (ii) appeared more common in summer than in winter.

The frequency used in these experiments was normally chosen in relation to ionospheric conditions to lie between $0 \cdot 8$ and $1 \cdot 2$ times the critical penetration frequency of the F_2 layer and the occurrence of the different scatter types is dependent upon the frequency used and its proximity to the F_2 layer critical condition. Consequently, if different frequencies had been used, slightly different results might have been obtained, e.g. if a sufficiently low frequency had been selected, types (i) and (ii) might have been more common. Alternatively, different results might have been obtained by selecting different observing times, e.g. night instead of day. Some observations continued into the night showed that, as the F_2 layer electron density decreased, the F_2 layer-dependent scatter types gradually moved out in time-delay range and decreased in elevation angle until they finally disappeared leaving only intermittent types (i) and (ii) echoes.

Usually, several backscatter types were simultaneously present and the scatter type combinations observed at Salisbury are listed in Table 1. Of these 17 recorded combinations, the four most commonly observed were, in decreasing order, (1) iv and v, (2) i and iv, (3) iv alone, and (4) i, iv, and v. All combinations involving types (iii) and (vi) were comparatively rare.

That other combinations were not observed does not deny their having occurred. It means only that they were not common and that they were not identified when they may have been present. The observed combinations are the simplest and they would be expected to be the most common. Two combinations which probably occurred and which escaped detection are (1) i, iv, vi and (2) i, iv, v, vi. These could have occurred without being recognized because of the extreme difficulty of simultaneously identifying the two essentially intermittent types (i) and (vi) in the presence of other types.

TABLE 1

OBSERVED SCATTER TYPE COMBINATIONS

Number	Combination	Number	Combination
1	i	10	i, iv, v
2	iv	11	iv, v, vi
3	i, ii	12	iv, v, vii
4	i, iv	13	i, iv, vii
5	iv, v	14	i, ii, iii, iv
6	iv, vi	15	i, ii, iv, v
7	iv, vii	16	i, iv, v, vii
8	i, ii, iii	17	i, ii, iii, iv, v
9	i, ii, iv		

IV. SCATTER TYPES (I), (II), (III), (VI), AND (VII)

While the principal purposes of these experiments at Salisbury were to compare backscatter from sea and from land and to detect scattering from prominent land masses using scatter types (iv) and (v), it was possible to make observations on other scatter types than those essential to these objectives.

Scatter types (vi) and (vii) were observed on a number of occasions and found almost exclusively between angles of 60 to 90° to the Earth's surface and principally from 75 to 90°. The elevation angle of type (vi) often varied through a wide range in a short time without appreciably altering in time delay, indicating the passage of sporadic-E clouds and the ionized trails of meteors. Type (vii) was most commonly observed when the experimental frequency was passing through the F_2 layer critical penetration frequency and it was seen to form at the trailing edge of the first F_2 vertical echo.

Types (ii) and (iii) were observed over a wide range of angles from 22 to 90°. Too few measurements were made on type (iii) to enable more to be said concerning it than that it appeared to be the rarest of the recorded types. Type (ii) occurred intermittently and usually persisted for a period of seconds to minutes although, on one occasion at least, it appeared on-and-off with remarkable persistence for over an hour. It has always been observed in the presence of type (i), but the angular ranges of these two types often differed. Type (ii) was more common in summer than in winter.

Type (i) was very common, having been identified during 65 of the 68 observational periods during summer and winter. It was more common during summer than in winter, though it would not appear so common during the summer midday period at Salisbury as in Europe. It occurred intermittently at all angles from 15 to 90° with widely varying field strengths. Sometimes, it persisted for long periods and, when it did, it showed rapid and deep fading—particularly in winter. Also, it often showed the large angle changes in a short time with no appreciable time-delay change which suggest the passage of sporadic-E clouds and the ionized trails of meteors. Some preliminary estimates of the speed of movement of the observed meteors indicated values between 30 and 50 km/sec.

Since the polar diagrams of the A120 and A300 aerials were so simple, an attempt was made to check the observed occurrence rate of type (i) echoes with elevation angle for these aerials against that expected if the scattering clouds of sporadic-E ionization were random in spatial occurrence for frequencies between 6·0 and 7·5 Mc/s. In accordance with this hypothesis, the number of echoes received in a given α-range should be proportional to the ionospheric area illuminated by the transmitted beam in this α-range. This would be reasonably true only near the vertical where the effects of signal attenuation and of the physical mechanism producing the scatter would not vary appreciably with angle. It was considered that this would hold down to 50°. The α-range, 51–90°, was subdivided into four divisions (51–60°, 61–70°, 71–80°, 81–90°) and the ratios of observed to expected occurrence rates were computed. The results obtained were in reasonable agreement with the stated hypothesis for the winter midday period at Salisbury. There were insufficient observations of the correct form to allow this comparison for summer. The winter ratios were :

(1) 0·84 for the α-range 51–60°
(2) 1·08 for the α-range 61–70°
(3) 0·84 for the α-range 71–80°
(4) 0·94 for the α-range 81–90°.

One effect of the presence of scatter types (i), (ii), and (iii) was to cause the amplitude of any simultaneously present types (iv) and (v) scatter to decrease and to fluctuate very rapidly. This was because part of the energy which normally would have produced scatter types (iv) and (v) had been used producing these other types. The presence of these three E region scatter types did not always influence the whole of the observed types (iv) and (v) echoes, e.g. if type (iv) was observed over the α-range 40–80° and a type (i) echo occurred with an α-range from 60–70°, it was only the type (iv) echo in the α-range of 60–70° which was affected. Sporadic-E ionization was sometimes sufficiently dense, particularly during summer, and persistent to prevent any energy reaching the F_2 layer and, consequently, it prevented the occurrence of the scatter types (iv) and (v).

V. SCATTER TYPES (IV) AND (V)

(a) *General Observations*

Summer observations on scatter types (iv) and (v) were made using only the A120 aerial and the more comprehensive winter data were obtained using the four transmitting aerials. The more intense and more persistent sporadic-E ionization in summer made observation of these two scatter types more difficult than in winter. Consequently, for these two reasons, the summer data are relatively incomplete compared to the winter data and the following discussion will refer only to the more comprehensive winter results. However, the summer and winter A120 measurements were in agreement on all points discussed below.

Most of the subsequent discussion, unless specifically stated otherwise, will refer to observing conditions when F_2 verticals were present. However, some comparison has been made between the scatter results when F_2 verticals were present and when absent. The principal observations were that, when no F_2 verticals were present :

(1) The α-range of the scatter was less, usually depending upon how far the frequency used was above the F_2 layer critical frequency.

(2) The scatter at the lower angles was stronger, i.e. below about 40°. This was undoubtedly due to the time-delay focusing of the rays at lower angles.

(3) As the F_2 layer electron density decreased, the types (iv) and (v) scatter, observed on a given frequency, moved out in time-delay range and decreased in elevation angle until type (v) disappeared and finally type (iv) disappeared.

It was normal to make observations using at least two of the four transmitting aerials on each day and type (iv) scatter was observed during all the winter observing periods. Type (v) was identified on 36 of these 40 days.

The angle of elevation frequency distributions for these two scatter types for the four differing transmitting aerials are shown in Figures 5–9 and these data form the principal basis of most of the subsequent discussions. These angle distribution curves give the total number of occasions or observations during all observing periods when scatter of a given type was observed at each elevation angle for each aerial when F_2 verticals were present. Observations were recorded at approximately 10-sec intervals and the angle was read with an accuracy of $\pm 1°$. The data sampling method was designed to produce comparable results for each scatter type on the four aerials. In the diagrams presented, no results are quoted for angles greater than 83° as, above this angle, the returned signals were not considered to be scatter but to be reasonably specular reflections.

The type (v) angle distribution characteristics were the same as those of type (iv), as was expected, with one exception. The lowest angle at which scatter was observed was lower for type (iv) than for type (v). Overall, type (iv) was observed at all angles from 22 to 83°. This low-angle discrimination was expected since scatter was generally weaker at the lowest angle and, in the case of multiple-hop scatter, this weaker low angle scatter would be the first to fall below the sensitivity of the receiving equipment.

It was also essential to understand the problems incurred by the use of four differing transmitting aerials before comparing the results obtained using them. The A-scope time delays and the CR.DF α-ranges observed at any given time were different on each of the four aerials. The three possible causes of this were examined and allowance was made for them when interpreting the experimental results, namely,

(1) the different aerial polar diagrams,
(2) the different scatter sources involved,
(3) the different ionospheric conditions with direction of transmission.

Additional complexity was introduced into the experimental technique by high-angle ray effects and ordinary/extraordinary (o/x) wave component effects.

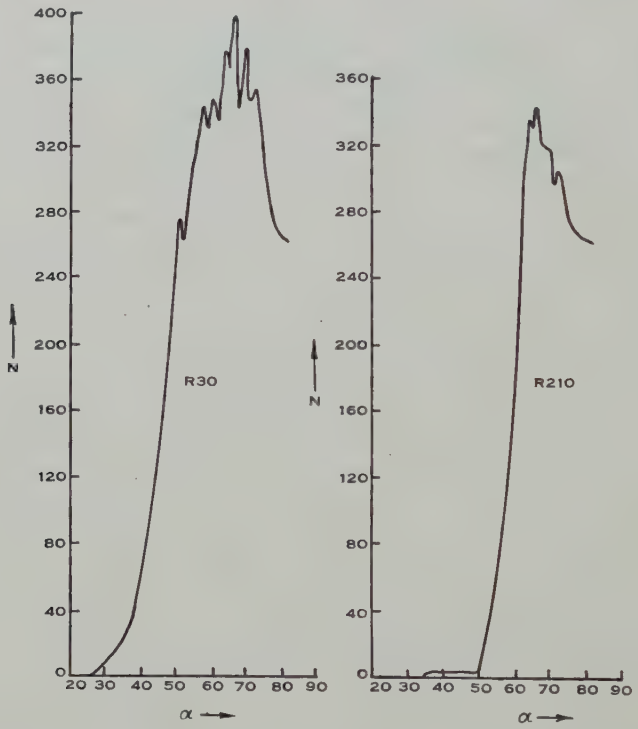


Fig. 5.—Number of occasions, N, during winter 1955 when scatter type (iv) was observed at various elevation angles (α) on the R30 and R210 aerials using scatter of all amplitudes.

High-angle ray effects were comparatively rare and were seen usually when the operating frequency was very close to the F_2 layer critical frequency and below it. One case of this effect was noted when the operating frequency was approximately $1 \cdot 2$ times the F_2 critical frequency. This effect was readily identified when present by the typical increase of elevation angle with time delay and was normally noted when only one wave component could be distinguished.

The o/x wave-component separation was observed frequently and, when it occurred, it was usually simple to identify the component producing the scatter. This was because the time-delay separation of the components was large enough, when associated with the time-delay focusing of the rays, to enable the o/x scatter components to be separated.

(b) *Scatter from Land and from Sea*

Backscatter from land and from sea were compared using the types (iv) and (v) elevation angle measurements made on the R30 and R210 aerials. The R30 aerial was directed over rough hilly ground towards Broken Hill and the R210 aerial was directed towards the Southern Ocean, i.e. mainly over sea. Two cases were examined, i.e. when F_2 verticals were present and when they were absent.

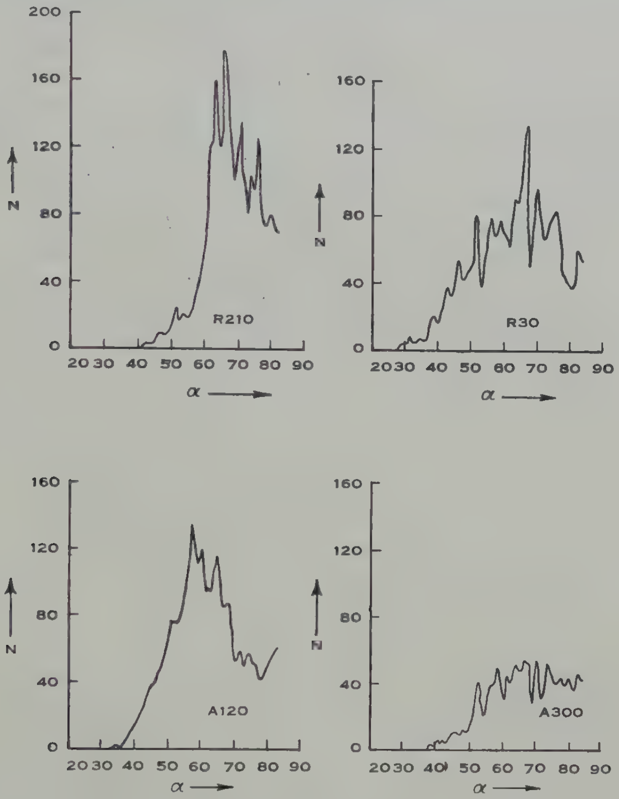


Fig. 6.—Number of occasions, N, during winter 1955 when scatter type (iv) was observed at various elevation angles (α) on the four aerials using only scatter amplitude maxima.

Considering the case when F_2 verticals were present, it can be seen from Figure 5 that, using backscatter of all amplitudes, the lowest angle at which type (iv) scatter was received using R30 was 29°, while, for R210, it was 43°. This was a difference of 14°. Similarly, from Figure 6, using only the observed amplitude maxima of type (iv) scatter, the lowest angle for R30 was 27° and, for R210, it was 35°—a difference of 8°. From Figures 7 and 8, it can be seen that for type (v) scatter, these differences are both 16°, with the R30 minimum being the lower in each case.

It was difficult to compare results when the operating frequency was greater than the F_2 critical frequency because there was no accurate method available at Salisbury for determining the F_2 critical frequency at any time. However, some comparison was possible for times when the frequency used was known to be very close to the F_2 critical frequency. Under these conditions, there was still a 10–12° difference between the lowest angles observed on the two R-aerials with R30's value the lower.

The general conclusion from the above observations is that the sea is not so pronounced a scatter source as is the land at low angles of elevation. This is certainly true for the present example. The complex polar diagram properties of the rhombics were examined and they do not influence this result, as the major side lobes were approximately in the forward direction, i.e. over land for R30 and over sea for R210.

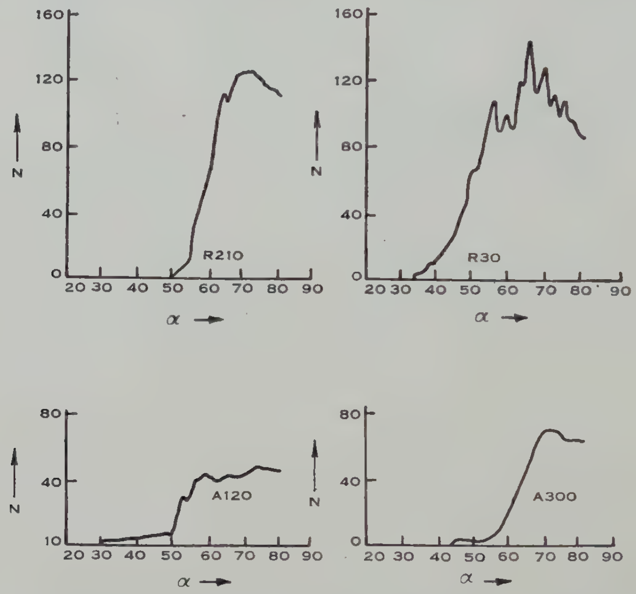


Fig. 7.—Number of occasions, N, during winter 1955 when scatter type (v) was observed at various elevation angles (α) on the four aerials using scatter of all amplitudes.

(c) *Scatter from Prominent Land Masses*

It was thought that the scattering effects of prominent land masses, such as mountain ranges, should be discernible in the A120 and A300 winter elevation angle data. Data, obtained only when F_2 verticals were present, were used to enable full use to be made of the broad polar diagram patterns of the array aerials in the elevation plane. These polar diagrams were measured to be reasonably uniform over the α-range from 22 to 90°. The data from R-aerials were not used in this study as they were used to demonstrate the land/sea difference and their polar diagrams were too complex to permit any detailed analysis of their corresponding angle data.

As for the land/sea comparison, the first point arises from an examination of the angle minima at which types (iv) and (v) scatter were received using the two A-aerials. It can be seen from Figure 9 that, using scatter of all amplitudes, type (iv) was received at 22° on A120 and the lowest angle for A300 was 31°. From Figure 6, using only the observed amplitude maxima, type (iv) was received at 34° on A120 and its lowest angle for A300 was 37°. It can be seen from Figure 7 that, using scatter of all amplitudes, type (v) was received at 31° for A120 and its lowest angle for A300 was 46°. Again, using only amplitude maxima, Figure 8 shows type (v) to have been received at 28° on A120 and at a minimum angle of 41° on A300. These observations suggest that there were more prominent land scattering sources at distances corresponding to low angles of elevation along the direction 120° E. of N. from Salisbury than along 300° E. of N. This was verified by the nature of the terrain as shown in Figure 4.

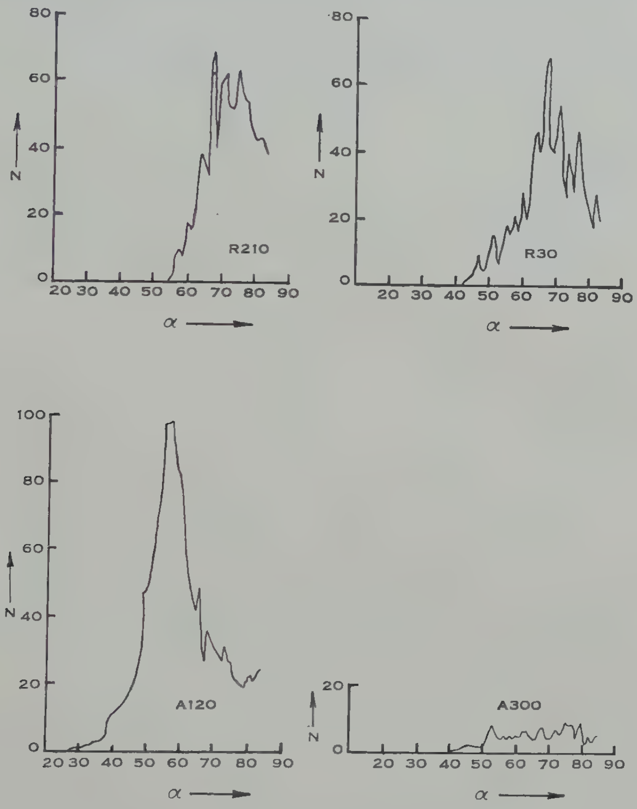


Fig. 8.—Number of occasions, N, during winter 1955 when scatter type (v) was observed at various elevation angles (α) on the four aerials using only scatter amplitude maxima.

As indicated previously, the backscatter echo pattern when viewed on the A-scope display, presented a ragged appearance with definite amplitude maxima. Angle measurements were possible on all echoes seen on the A-display independently of their amplitude or only on the amplitude maxima echoes. The A120 and A300 type (iv) angle distributions shown in Figure 9 were examined. The A120 measurements show a maximum occurrence rate at 58° and there tends to be another maximum towards 90°. The A300 data show an occurrence maximum between 65 and 74° and there again tends to be another maximum towards 90°. These 90° maxima were due to the occurrence of the second-order F_2 vertical reflections, but the other maxima must be due to the presence of prominent scattering sources on the ground at distances corresponding to these angles. The possibility that these maxima could have been due to receiver aerial characteristics was examined and a negative result was obtained. The same occurrence maxima can be observed if the type (iv) amplitude maxima data are used, as seen in Figure 6. Figures 7 and 8 show that the type (v) observations present this same result.

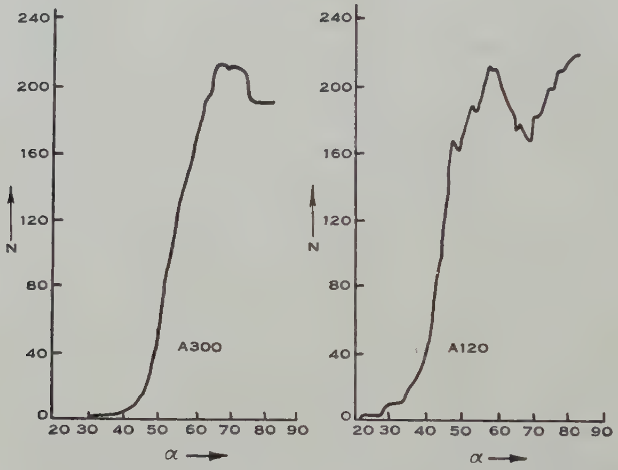


Fig. 9.—Number of occasions, N, during winter 1955 when scatter type (iv) was observed at various elevation angles (α) on the A300 and A120 aerials using scatter of all amplitudes.

A search for the possible scatter sources which would produce these occurrence and amplitude maxima revealed that an excellent correspondence was obtained if the A120 scatter maxima at 58° were caused by the Grampian mountains at the end of the Great Dividing Range and if the A300 scatter maxima between 65 and 74° were caused by the Gawler Range—see Figure 4.

The possibility that these occurrence and amplitude maxima were merely apparent and due to the blanketing effects of sporadic-E ionization was also examined and found to be unimportant. It was concluded that, if any sporadic-E blanketing effect did occur, it would tend to minimize the effects described. Also, the recurrence maxima at 68° for type (iv) scatter on the R-aerials were not significant as they were due to subsidiary maxima in the polar diagrams of these aerials.

VI. CANBERRA AND SALISBURY F_2 LAYER CRITICAL FREQUENCIES

There is an ionospheric observatory at Canberra which is approximately 1000 km due east of Salisbury. It was thought possible that the Canberra vertical incidence data would be suitable for comparison with the Salisbury backscatter data. However, it was found that the F_2 layer critical penetration frequencies measured at Canberra were nearly always less than those indicated at Salisbury by the presence of F_2 layer vertical returns on various frequencies when adjusted to the same L.M.T.'s.

On about 30 occasions during the summer midday period when a comparison was possible, the Canberra F_2 critical frequency was of the order of 1 Mc/s less than the indicated Salisbury value. On about 70 occasions during the winter midday period when a comparison was possible, the Canberra F_2 critical frequency was of the order of $0 \cdot 6$ Mc/s less than the indicated Salisbury value. The magnitude of this difference, which was not expected, prevented the suggested comparison. This observation may be of particular interest to users of ionospheric communication circuits.

VII. CONCLUDING REMARKS

High-frequency backscatter was examined experimentally at Salisbury during summer 1954 and during winter 1955 using a pulse technique which consisted of simultaneously measuring the elevation angle of arrival and the time delay (with respect to a ground pulse of zero delay) of the scatter. Observations were confined mainly to conditions when the frequency used was between $0 \cdot 8$ and $1 \cdot 0$ times the F_2 layer critical penetration frequency.

Seven backscatter types were distinguished which showed both ionospheric and ground origins. The most common of these was backscatter from the ground via the F_2 layer—type (iv). Combinations of backscatter types were observed and the most complex of these was a combination of types (i), (ii), (iii), (iv), and (v). A comparison of scatter from land and from sea revealed the land to be the more prominent scatter source at low angles of elevation. Further, scatter from outstanding land masses was identified during the observations on types (iv) and (v).

The effects of aerial polar diagrams on backscatter were examined and found to be important. High-angle ray and ordinary/extraordinary wave component effects were studied as also was the influence of the occurrence of scatter types (i), (ii), and (iii) on the amplitude and occurrence rates of types (iv) and (v).

An unexpected result, found in connexion with this backscatter investigation, was that the F_2 layer critical penetration frequency at Canberra was almost invariably significantly lower during the midday period than that at Salisbury.

VIII. ACKNOWLEDGMENTS

The aircraft used for the aerial polar diagram measurements and for the angle of elevation calibration was provided by the Aircraft Research and Development Unit of the Royal Australian Air Force.

This work was carried out by the author under the supervision of Mr. C. G. Carter of the Weapons Research Establishment, and the permission of the Department of Supply to publish the results is acknowledged.

IX. REFERENCES

ABEL, W. G., and EDWARDS, L. C. (1951).—*Proc. Inst. Radio Engrs., N.Y.* **39** : 1538.
BENNER, A. H. (1949).—*Proc. Inst. Radio Engrs., N.Y.* **37** : 44.
DIEMINGER, W. (1951).—*Proc. Phys. Soc.* B **64** : 142.
ECKERSLEY, T. L. (1940).—*J. Instn. Elect. Engrs.* **86** : 548.
ECKERSLEY, T. L., MILLINGTON, G., and COX, J. W. (1944).—*Nature* **153** : 341.
EDWARDS, C. F., and JANSKY, K. G. (1941).—*Proc. Inst. Radio Engrs., N.Y.* **29** : 322.
HARTSFIELD, S. M., OSTROW, S. M., and SILBERSTEIN, R. (1950).—*J. Res. Nat. Bur. Stand.* **44** : 199.
HARTSFIELD, S. M., and SILBERSTEIN, R. (1952).—*Proc. Inst. Radio Engrs., N.Y.* **40** : 1700.
KONO, T. (1950).—Rep. Iono. Res. Japan **4** : 127.
McCUE, C. G. (1955).—Dept. Supply, Aust., Weapons Research Establishment, Rep. Nos. WRE/ERD. 28 and 33.
PETERSON, A. M. (1949).—*J. Geophys. Res.* **54** : 284.
PETERSON, A. M. (1951).—*J. Geophys. Res.* **56** : 221.
SILBERSTEIN, R. (1953).—*Science* **118** : 759.
SILBERSTEIN, R. (1954).—*Trans. Inst. Radio Engrs., N.Y.* **AP-2** : 56.
DE WALDEN, S., ROCKE, A., BARRETT, J., and PITTS, W. (1947).—*J. Instn. Elect. Engrs.* IIIA **94** : 823.

RADIO ECHO OBSERVATIONS OF METEOR ACTIVITY IN THE SOUTHERN HEMISPHERE

By C. D. ELLYETT* and C. S. L. KEAY*

[*Manuscript received August* 31, 1956]

Summary

A further analysis has been made of 32,000 meteor echoes obtained during 1953 by radar apparatus operating at a frequency of 69 Mc/s. Monthly and annual mean diurnal rate curves have been drawn, revealing a strong similarity between southern and northern hemisphere results. It is concluded that most meteoric matter incident on the southern hemisphere, down to magnitude +4·5, is confined to direct orbits closely following the plane of the ecliptic.

Subsequently the sensitivity of the equipment was increased to yield echo rates much higher than the equivalent visual rates. Analysis of the results obtained during February and March 1956, which are normally regarded as being months devoid of showers, reveals discrete radiant activity near the helion and antihelion positions. This result largely removes the former distinction between shower and sporadic meteors, since many of the meteors previously regarded as sporadic represent the upper limit of showers of minor-sized particles.

I. INTRODUCTION

A survey of meteor activity in the southern hemisphere was carried out during 1953, using conventional radar methods (Ellyett and Roth 1955). Radiants were calculated by the Clegg technique (1948), from range-time plots of meteor echoes. A rotatable beam aerial was used, directed respectively 22½ °N. and 22½ °S. of W. on alternate days. These radiant results, together with the known northern hemisphere shower data, showed that the major percentage of all showers, from either hemisphere, lay fairly close to the plane of the ecliptic. The southern results, representing meteors down to a magnitude of +4·5, have now been further analysed in terms of monthly and annual mean diurnal rates, and show that most of the activity in the southern hemisphere arises from particles having direct orbital motion of low inclination.

Following this work, a departure has been made from the usual technique (Lovell, Banwell, and Clegg 1946 ; Hawkins and Almond 1952 ; and others) of maintaining the radar rate closely similar to the visual rate. The visual rate is essentially a property of the eye rather than of the total incident meteor flux. The sensitivity of the present radar equipment, operating at 69 Mc/s, has therefore been increased in order to record a much higher rate of specular reflections from small meteors. Approximate radiants, determined from the total rates, have shown that meteors appearing as sporadic at sensitivities equivalent to visual become part of a definite shower distribution at a higher

---

* Canterbury University College, Christchurch, New Zealand.

sensitivity. This largely removes the former distinction between shower and sporadic meteors, since most of the meteors previously regarded as sporadic represent the upper limit of showers of minor-sized particles.

II. Monthly and Annual Mean Diurnal Rates

The aerial system and radar equipment used for the 1953 southern hemisphere meteor survey have already been fully described (Ellyett and Roth 1955). It is therefore sufficient to state that pulses of 75 kW peak power and 3·5 μsec duration were radiated at a frequency of 69 Mc/s and a recurrence rate of 145/sec. The aerial power gain was 75 compared with a half-wave dipole ; and the receiver, of bandwidth 300 kc/s, had a noise factor of 6·8 dB.

A sporadic meteor has been defined as one which cannot be associated with any definite radiant. Monthly mean diurnal curves of sporadic meteor rate at 72 Mc/s have been given by Lovell (1954) for the northern hemisphere, over a 2-year period from October 1949 to September 1951.*

Owing to the complexity and degree of interlocking of the southern hemisphere night meteor showers in the June-August period and the still doubtful nature of a number of the more minor radiants, it has not proved possible in this case to separate the showers from the sporadic background. Groupings of longer-range meteors stand out clearly on range-time plots, and provide identification of most showers, but at the lower ranges such groupings overlap and mix randomly with sporadic meteors. Attempts to remove the meteors associated with any particular radiant become arbitrary and subjective. It has, however, gradually become recognized that shower meteors, at the visual observing level, constitute but a small proportion of the total influx. This has been noted for visual showers by Öpik (1934) and by Levin (1955). Hawkins (1956*a*), using radar equipment at a frequency of 72 Mc/s and a sensitivity similar to the present survey, finds that showers form less than 5 per cent. of the total number of incident meteors. The shape of the monthly mean rate curves at approximately the visual rate level should therefore not be altered greatly by including meteors belonging to showers. Such curves for 43·5 °S. latitude are given in Figure 1.

Equipment failure occurred in May 1953. Consequently the graphs for May 1954, obtained at the same equipment sensitivity, have been included, but have not been used in further analyses.

A strong general similarity is immediately evident between northern and southern hemisphere activity. In particular both groups show minimum activity in February, a progressive rise to June, almost comparable activity in July, and an intermediate rate in the second half of the year. The similarity of activity in the two hemispheres removes any possibility of appreciable ionospheric control of the recorded meteor rate.

One point of difference is in the May, June, July activity, which is greatest in the day-time hours in the northern hemisphere and during the night hours in the southern hemisphere. The northern daylight streams do not reach a high elevation at 43·5 °S. and are overshadowed by the strong southern night radiants which occur at a mean declination of about $\delta=-20°$.

---

* See also Hawkins (1956*b*).

Following Lovell (1954), the observations shown in Figure 1, representing 32,000 echoes, although in some months not as complete as desired, have been used to obtain the annual mean curve for the diurnal variation of meteor rate. This curve is given in Figure 2 (*a*), from which Figure 2 (*b*) was obtained by normalizing, or giving equal weight to each month. Marked maxima occur in the curves at mean times of $01^h\,25^m$ and $10^h\,10^m$ L.T. In this respect the result is almost identical with that obtained in the northern hemisphere. Both the 1949–50 northern survey and the 1953 southern survey show a similar lack of symmetry of the maxima about the apex of the Earth's way.

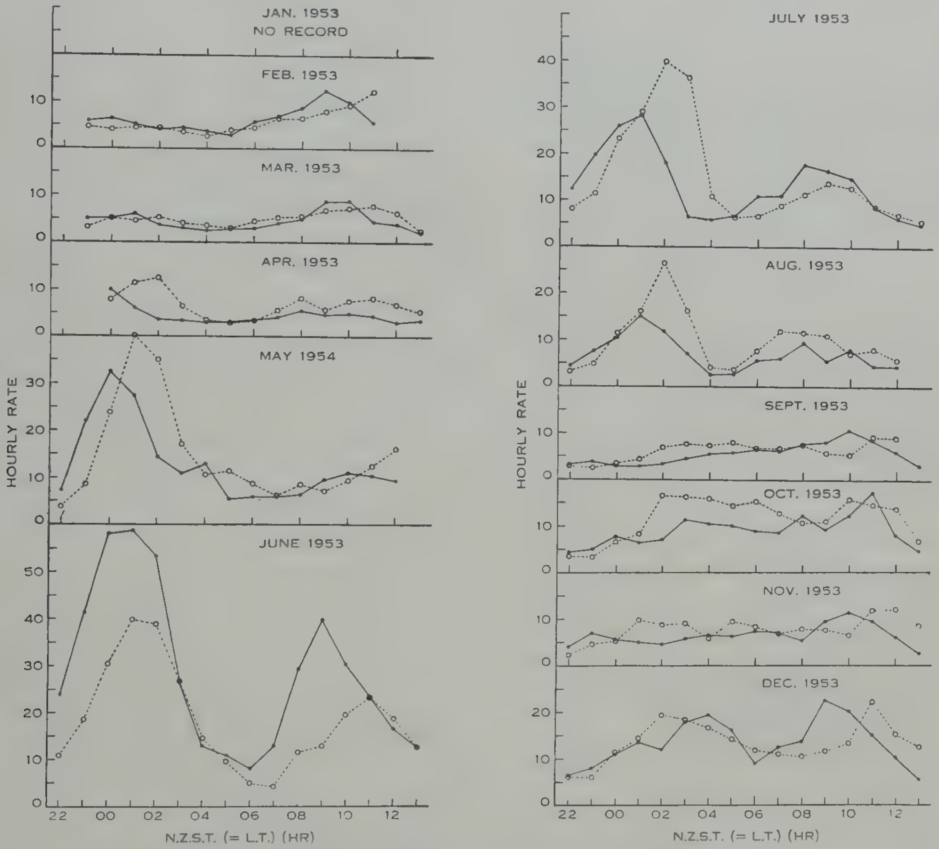


Fig. 1.—Monthly mean diurnal rates of meteor activity.
○ - - - ○ Southerly aerial. ●——● Northerly aerial.

The northern maxima have been recognized (Almond, Davies, and Lovell 1952, 1953; Hawkins 1956*a*) as being due to meteors from near the helion and the antihelion directions, and hence represent a group of meteors moving in direct low inclination low eccentricity orbits. A surprising feature of the southern results is the almost complete absence of a peak at 0600 hr L.T.

Consequently, since the aerial polar diagram was similar to that used by Hawkins, and since showers have been included, the conclusion must be that the greater part of *all* meteoric matter incident on the southern hemisphere, down to the survey magnitude of $+4\cdot5$, is confined to direct orbits closely following the plane of the ecliptic.

This conclusion receives strong support from the photographic work of Whipple (1954), where both stream and sporadic meteors in the northern hemisphere are shown to be highly concentrated within an inclination of 35°. Kresák (1955) has reached a similar conclusion from a study of northern telescopic meteors. There are of course a number of very well-known high inclination northern showers. The possibility of similar southern showers cannot yet be ruled out, as surveys are incomplete. Likewise high inclination northern sporadic meteors are observed (Whipple 1951, 1954), which again may have their counterpart in the south. Nevertheless, the greater proportion of both sporadic and shower meteors in the southern hemisphere appear to be ecliptical.

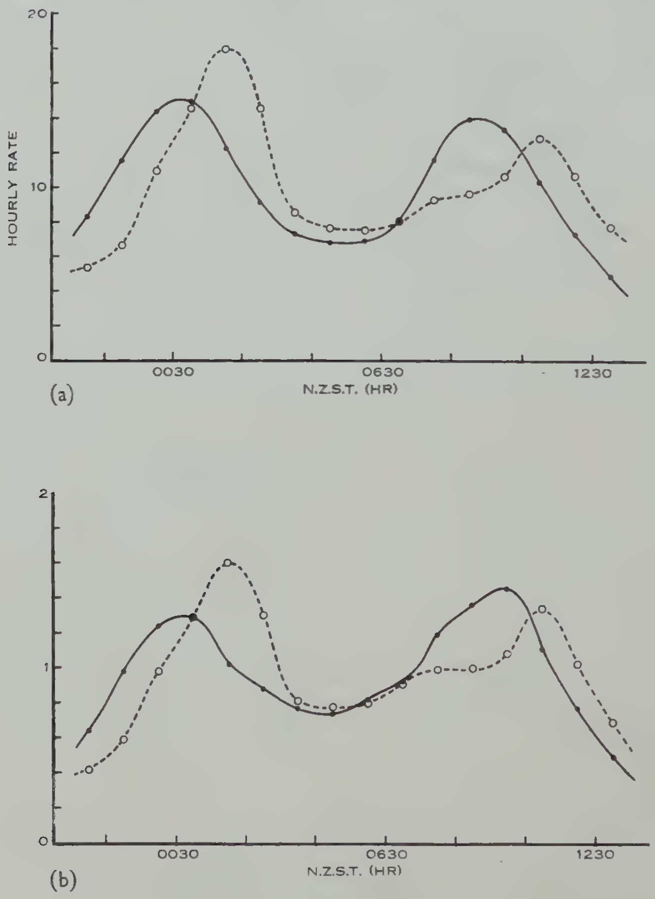


Fig. 2 (*a*).—Annual mean diurnal rate of meteor activity.
Fig. 2 (*b*).—Normalized annual mean diurnal rate.
○ - - - ○ Southerly aerial. ●——● Northerly aerial.

The only other known estimate of southern hemisphere radar rates, in this case for a mixture of sporadic meteors and minor showers, was obtained by Weiss (1955). His theoretical calculations on background meteors are based on a uniform distribution of radiants and activities. All the experimental evidence described above indicates an anisotropic distribution : hence theoretical rate curves calculated on an isotropic basis are no longer valid. Insufficient experimental data have been published by Weiss to make possible any significant comparison with the present results.

III. Showers of Minor Particle Size

(*a*) *Experimental Method*

Following a study of the requirements for maximum echo rate detection (Ellyett and Fraser 1955), it was decided to realize more fully the capabilities of the present radar equipment. This involved the sacrifice of all parity with visual rates. High-rate experiments were carried out during February and March 1956 ; the rate being achieved by increasing the pulse width from 3·5 to 26 μsec. The receiver noise figure was kept at the relatively poor value of 6·8 dB to ensure that cosmic noise variations did not influence the meteor rate.

These two months were chosen to correspond with the period of minimum meteor activity. The northern curves of monthly mean diurnal activity (Lovell 1954) show very little sign of the helion and antihelion maxima during these months ; and no appreciable showers occur. The 1953 southern curves do give a definite but small indication of a mid-morning activity. McIntosh (1935) records a very weak visual Corona Australis shower in mid March, but otherwise there are no significant showers known for this period in the southern hemisphere.

Rates during February were maintained at about 1000 to 2000 echoes in each daily 16-hr observing period. This rate corresponds to a limiting magnitude of +7·5. Because of the known deficiency of large meteors throughout this period of the year, the limiting magnitude may be above +7·5, since this value has been calculated by a rate comparison method. In the first 6 days of March, and after the 14th day an intermediate rate of about 625 echoes per 16 hr was maintained. From March 7 to 14 the rate was held at the low value of 175 per 16 hr, which is comparable with the 1953 survey rate.

Construction of high-rate range-time plots from the echoes on the photographic film becomes impossibly time-consuming. An example of such a plot is given in Figure 3. Drawing the usual vertical line for each meteor would completely obscure all the lower range echoes. It is difficult, by inspection of plots such as Figure 3, to find radiant groupings.

The method which was adopted for both the intermediate and the high rates is shown for the intermediate-rate plot of Figure 4 (*a*). Such plots were not normally made, but only the rate was read from the film directly, over half-hourly periods, to give the result shown in Figure 4 (*b*). Owing to the random nature of the occurrence of meteor echoes, even within a shower group, there will be statistical fluctuations in the echo rate. Consequently the half-hourly rate curve was smoothed in groups of three consecutive values, the central value being weighted by a factor of 2. Since the minimum width of a significant radiant peak is of the order of $1\cdot0$–$1\cdot5$ hr, it was considered unwise to smooth over more than three values.

Approximate radiants can be obtained from families of these daily curves. (A simple method of obtaining accurate radiants from high-rate plots has since been devised, and forms the basis of a subsequent paper.)

(*b*) *Results*

The rate curves obtained, of which a selection is given in Figure 5, reveal a mid morning peak of activity persisting right through February, and still apparent, although shifted slightly in position, during most of March. The time displacement of the peak between the records from the two aerial positions was constant enough to prove that the cause was neither instrumental nor terrestrial. Some prominent but less regular peaks were also present between 0000 and 0400 hr L.T.

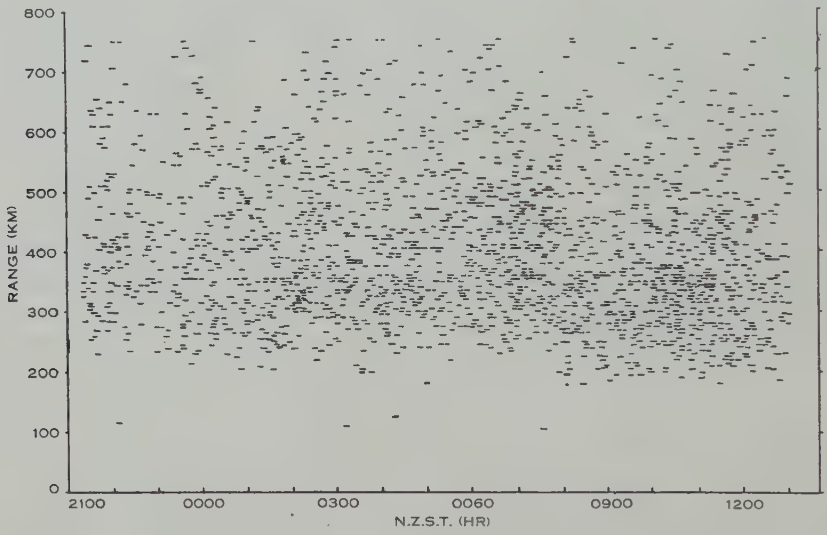
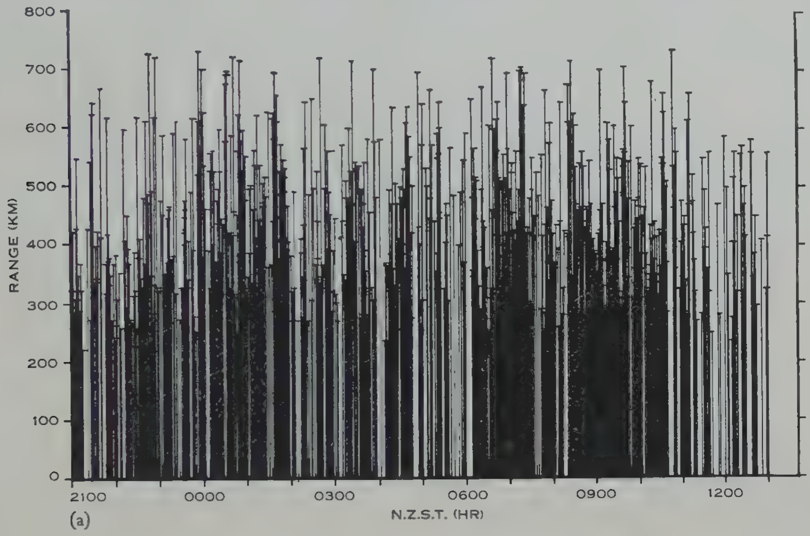


Fig. 3.—Range-time plot, February 4–5, 1956. Aerial azimuth 292½°.

In order to find the coordinates of the radiant from which the mid-morning activity originated, assuming that there was only one radiant, the times of occurrence of the maxima were tabulated in weekly groups after reduction to a datum day. The declination was found from the average time difference between the maxima in each aerial. Range-time plots, adapted for the southern hemisphere from Clegg (1948), indicated that the time of local transit would occur approximately 1 hr after the mean of the times of maxima in the two aerials. In this way the Right Ascension was found. With this method of analysis, the maximum error should not exceed ±5° in declination, and ±10° in Right Ascension. The results are given in Figure 6, together with radiant positions of definite southern showers which occur at other months of the year (Ellyett and Roth 1955).

The newly found activity, of which some hint was given in the 1953 monthly diurnal curves, is again close to the plane of the ecliptic. It seems probable that there are either one or two fairly diffuse continuing radiants present which have a tendency to shift with the Sun. The diffuseness of the radiant is indicated by the breadth of the peak on the rate curve.



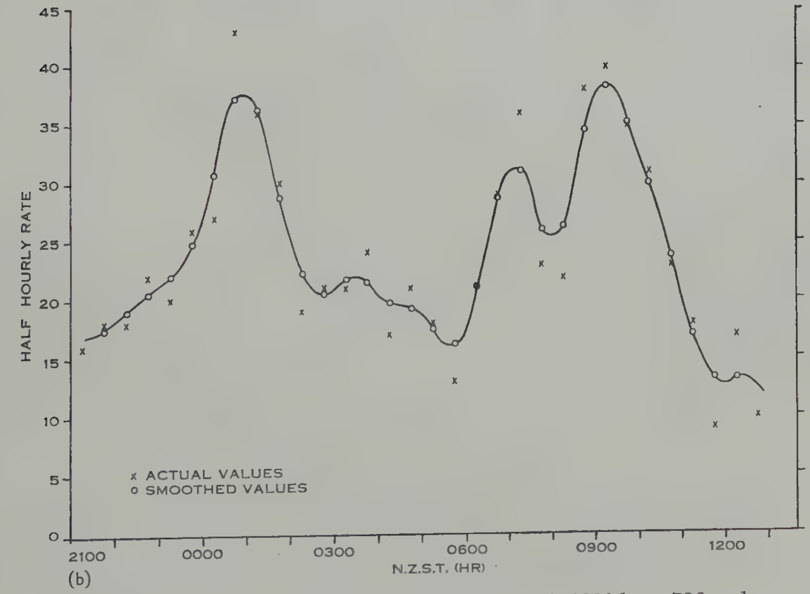


Fig. 4 (*a*).—Range-time plot, March 21–22, 1956, 2130–1230 hr. 736 echoes. Aerial azimuth 292½°.

Fig. 4 (*b*).—Smoothed half-hourly rates, March 21–22, 1956. Aerial azimuth 292½°.

A major fact emerging from these results is that small, random-looking groupings on the low-rate curves, which in the past have been regarded as due to sporadic meteors, build up, at the higher rates, into definite peaks with all the characteristics of meteor showers. This can be seen by comparing the rate curves of Figure 5. These new showers agree with low inclination direct motion groupings of meteoric matter and contain very few particles large enough to be detected by radar of low sensitivity. Such a type of shower might well be expected from the mass distributions found in the major northern showers. Generally, but not always, these showers are rich in bright meteors (Millman 1954). One exception is the day-time northern Arietid shower (Kaiser 1953) which, by its paucity of large meteors, conforms in type to the present February-March activity. Liller (1949) finds some of the northern showers are accom-

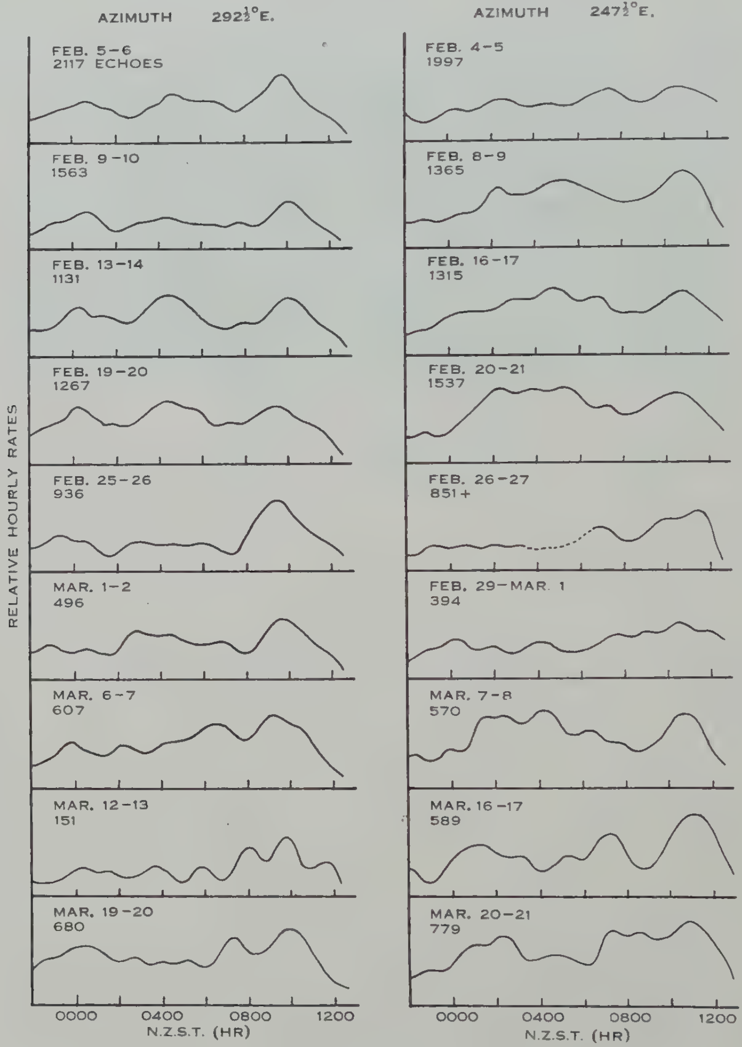


Fig. 5.—Meteor rate curves (smoothed).

panied by numerous bursts of sporadic E at 3·5 Mc/s, while others show no bursts, and some burst maxima occur without a shower.

Again, many workers have pointed to the similarity in composition of shower and sporadic meteors (Porter 1944; Kaiser 1953). The present work would indicate that sporadic meteors must acquire a much more limited definition.

(c) *The λ Corona Australid Shower*

This shower is a rather small one, recorded visually by McIntosh (1935); and detected on radar equipment by Weiss (1955), who records a peak of activity on March 15–16. The shower is of fairly high declination (−40°/−50°), and culminates at approximately 0600 hr L.T. By a comparison of simultaneous rate curves on equipments of different sensitivities Weiss concludes that this shower is quite deficient in large meteors.

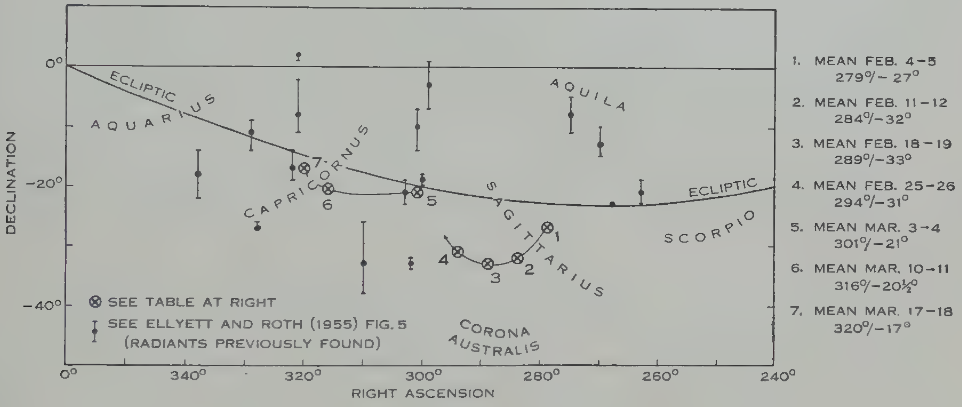


Fig. 6.—Plot and table of minor shower radiants found during February-March 1956.

During March 1956 no evidence was found for this shower. The equipment sensitivity from March 8 to 14 gave an echo rate of about 10/hr, which is equivalent to the sensitivity used by Weiss. The higher sensitivity used from March 15 to 23, giving a rate of 50 echoes/hr, also failed to reveal the shower. From these results it can only be concluded that the λ Corona Australid shower is variable in activity from year to year.

IV. Acknowledgments

This paper represents part of the New Zealand programme of meteor research now financed jointly by the University of New Zealand and the New Zealand Department of Scientific and Industrial Research.

V. References

Almond, M., Davies, J. G., and Lovell, A. C. B. (1952).—*Mon. Not. R. Astr. Soc.* **112**: 1, 30.

Almond, M., Davies, J. G., and Lovell, A. C. B. (1953).—*Mon. Not. R. Astr. Soc.* **113**: 4, 425.

Clegg, J. A. (1948).—*Phil. Mag.* **39**: 577.

Ellyett, C. D., and Fraser, G. J. (1955).—*Aust. J. Phys.* **8**: 273.

ELLYETT, C. D., and ROTH, K. W. (1955).—*Aust. J. Phys.* **8**: 390.

HAWKINS, G. S. (1956*a*).—Inter. Rep. Harv. Astr. Obs. No. 12.

HAWKINS, G. S. (1956*b*).—A radio echo survey of sporadic meteor radiants. *Mon. Not. R. Astr. Soc.* (in press).

HAWKINS, G. S., and ALMOND, M. (1952).—*Mon. Not. R. Astr. Soc.* **112**: 219.

KAISER, T. R. (1953).—*Advanc. Phys.* **2**: 528.

KRESÁK, Ľ. (1955).—"Meteors." (Ed. T. R. Kaiser.) p. 161. (Pergamon Press: London.)

LEVIN, B. J. (1955).—"Meteors." (Ed. T. R. Kaiser.) p. 141. (Pergamon Press: London.)

LILLER, N. (1949).—Tech. Rep. Cruft Lab., Harvard Univ. No. 65.

LOVELL, A. C. B. (1954).—"Meteor Astronomy." pp. 112–6. (Clarendon Press: Oxford.)

LOVELL, A. C. B., BANWELL, C. J., and CLEGG, J. A. (1946).—*Mon. Not. R. Astr. Soc.* **107**: 164.

MCINTOSH, R. A. (1935).—*Mon. Not. R. Astr. Soc.* **95**: 709.

MILLMAN, P. M. (1954).—*Science* **120**: 325.

ÖPIK, E. J. (1934).—Circ. Harv. Astr. Obs. No. 388.

PORTER, J. G. (1944).—*Mon. Not. R. Astr. Soc.* **104**: 265.

WEISS, A. A. (1955).—*Aust. J. Phys.* **8**: 148.

WHIPPLE, F. L. (1951).—"Compendium of Meteorology." p. 356. (Amer. Met. Soc.)

WHIPPLE, F. L. (1954).—*Astr. J.* **59**: 212.

A SURVEY OF COSMIC RADIO EMISSION AT 600 Mc/s

By J. H. Piddington* and G. H. Trent*

[*Manuscript received June 1, 1956*]

Summary

The results are given of a survey of cosmic radio emission at 600 Mc/s between declinations 90 °S. and 51 °N. using a 3°·3 wide beam. Isophotes are plotted in a system of celestial coordinates (epoch 1955).

The determination of radio spectra from surveys at different frequencies is discussed. A vital factor which tends to be neglected is the allowance made for side and back aerial lobes ; a method of making such allowance is suggested.

The possible sources of the emission are discussed. Away from the galactic plane the radiation is non-thermal. On the galactic plane and near the galactic centre the evidence, while not conclusive, suggests a mainly thermal origin.

The radio isophotes are used to determine a galactic pole. In Lund coordinates this is at $l \sim 330°$, $b = 89°\cdot 1$ with the Sun at a distance of about 50 parsecs north of the galactic plane.

I. Introduction

A survey of cosmic radio emission at 600 Mc/s has been made over most of the celestial sphere between declinations 90 °S. and 51 °N. from a field station near Sydney, N.S.W. In a previous communication (Piddington and Trent 1956) the localized sources found during the survey were discussed. In the present paper the general background of radiation is described.

Radio surveys of the whole southern Milky Way are available at 200 Mc/s (Allen and Gum 1950) using a beam of width 25°, and 100 Mc/s (Bolton and Westfold 1950) using a beam of width 17°. Attempts were made to deduce from these surveys the true brightness distribution by making allowance for the blurring effect of the wide aerial beam. It is now known that detail lost because of finite beamwidth cannot be recovered by subsequent graphical or other processes (Bracewell and Roberts 1954) and that the 100 Mc/s and 200 Mc/s plots of brightness distribution in the region of the galactic plane cannot be accepted as even approximately accurate. No full survey has been made with a narrower beam.

The present survey was made with a beam of width 3°·3 between half-power points. No attempt is made to reconstruct the observed brightness distribution plot to correct for the aerial beamwidth. Any structure smaller in size than this width is lost.

The aerial system, radiometer, and calibration equipment have been described in the previous (discrete source) paper and will not be discussed further. The observations were made by setting the aerial at a fixed declination

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Chippendale, N.S.W.

and allowing the emitting regions to pass through the stationary beam. The declination was then changed by $1\frac{1}{2}$° and the process repeated. Several repetitions of some plots were made to eliminate the effects of external interference and random drifts in the equipment. All plots were repeated at least once.

II. Results

The experimental results were reduced in the usual manner to a system of isophotes giving the aerial beam temperature* for any direction. The temperatures† are not absolute but above a minimum value, which persists in the cold parts of the sky, and which is discussed in Section III. The plot is shown in Figure 1 (Fig. 1 (*a*) shows the full plot, Figs. 1 (*b*), 1 (*c*), and 1 (*d*) enlarged sections) in celestial coordinates (epoch 1955). To allow estimates to be made of the different degrees of " blurring " which might be expected in different parts of the plot the aerial response diagram is included. It should be noted that when the beam is swung north or south of the celestial equator the width in minutes of Right Ascension increases as the secant of the declination angle.

Discrete sources which lie within or near a contour line have been retained in the plot. This is desirable since their separation from the background is often a matter of judgement. The strongest sources, such as those in Cygnus near Dec. 40 °N., are enclosed by several contour lines. The weakest sources do not raise the intensity by even one contour interval ; these are outlined by dotted lines. Several examples are seen near R.A. 12^{h}, Dec. 60 °S. Isolated sources, lying outside the minimum intensity contour, are not shown but a full list of all sources has been given in an earlier paper (Piddington and Trent 1956).

The general pattern of brightness distribution is similar to those in earlier surveys. It closely follows the Milky Way, the lowest measurable level of brightness extending out to 40° latitude but generally being confined to 25° or less. The outstanding features of different surveys change with frequency and beamwidth and in this survey they comprise the " galactic nucleus " (17S2A), the sources in Cygnus (19N4A and 20N4A), the main Centaurus source (13S4A), and the main Taurus source (05N2A) in that order followed by a string of sources lying nearly on the galactic plane and probably comprising thermally emitting H II regions. The elongation of the sources near Dec. 60 °S. is due to distortion by the linear plot used.

The plot may be compared with that of Reber (1948) on the neighbouring frequency of 480 Mc/s. A common feature of both is the small amount of radiation received from the general direction of the anticentre where, at 600 Mc/s, there are large parts of the galactic plane yielding beam temperatures less than 5 °K above the background. A more detailed comparison reveals a difference between the two plots in this region (about R.A. 06^{h}). Reber's emission region

---

* The precise meaning of this term and its implications are described in Section III.

† The absolute temperatures may readily be obtained from Figure 1 by applying the corrections given at the end of Section III.

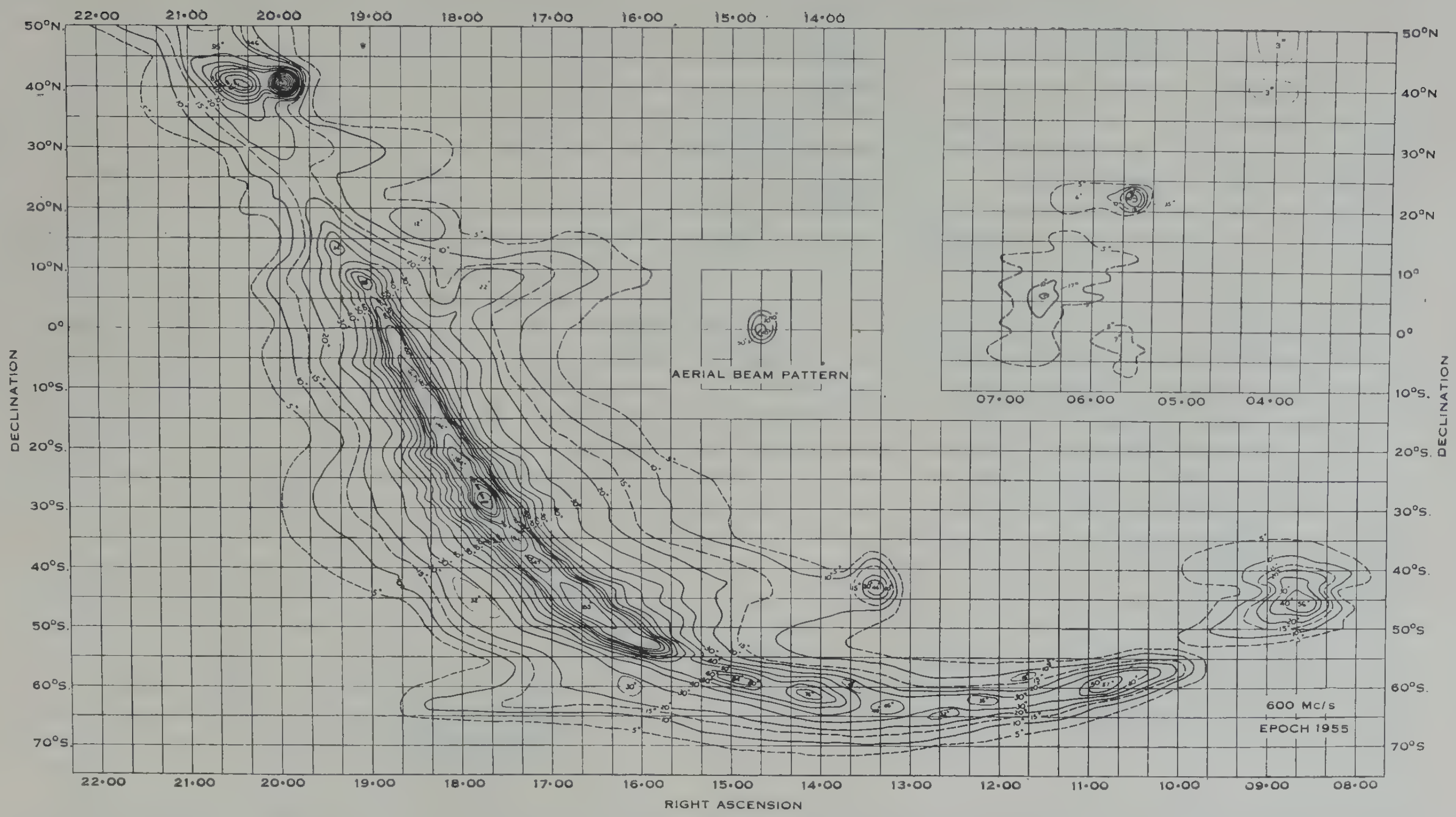


Fig. 1 (*a*).—Isophotes of cosmic radio emission. The quantity plotted is " beam temperature " which is defined as 100/65 times the observed effective aerial temperature. The full plot.

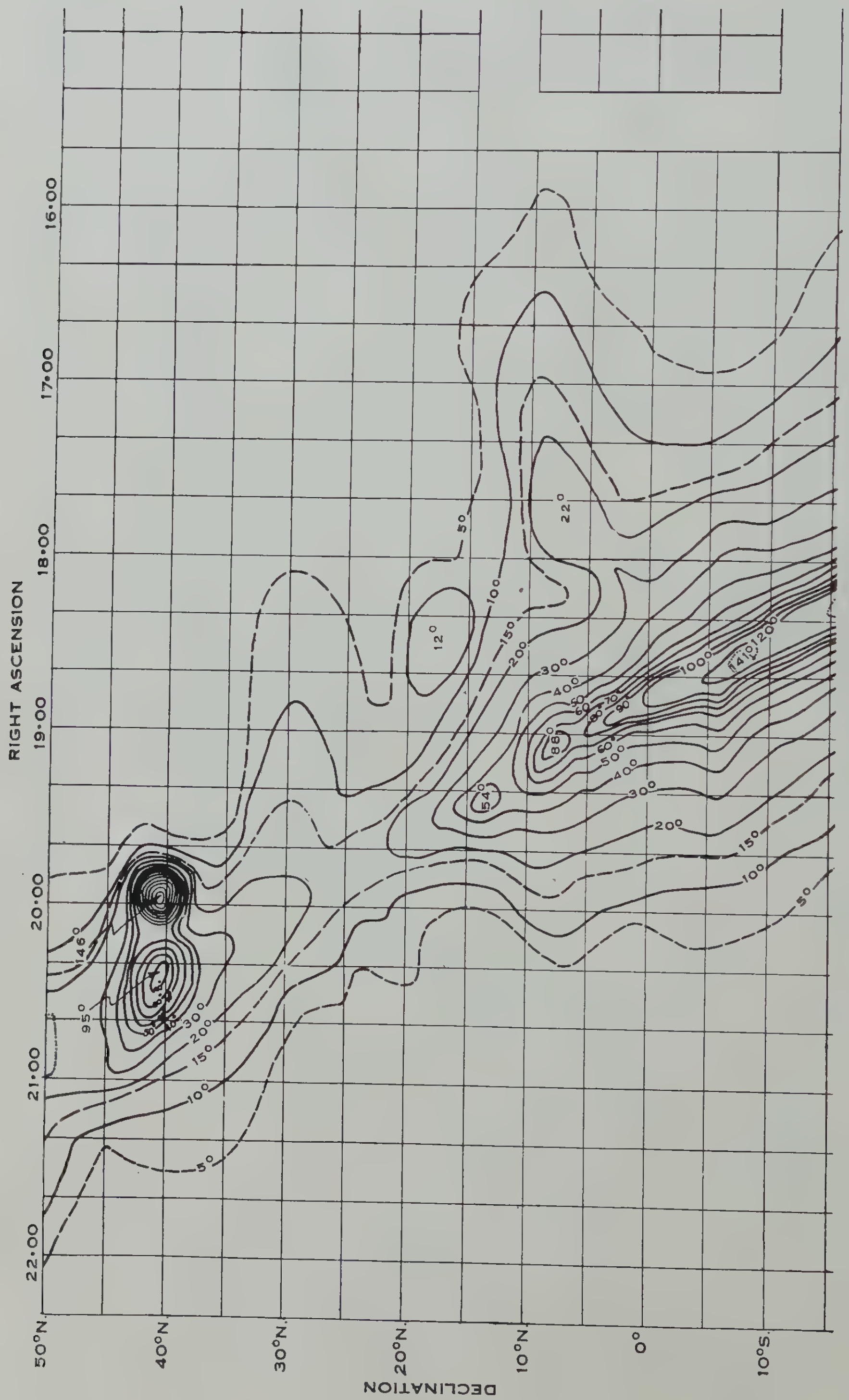


Fig. 1 (*b*).—Isophotes of cosmic radio emission. The plot of the northern sky.

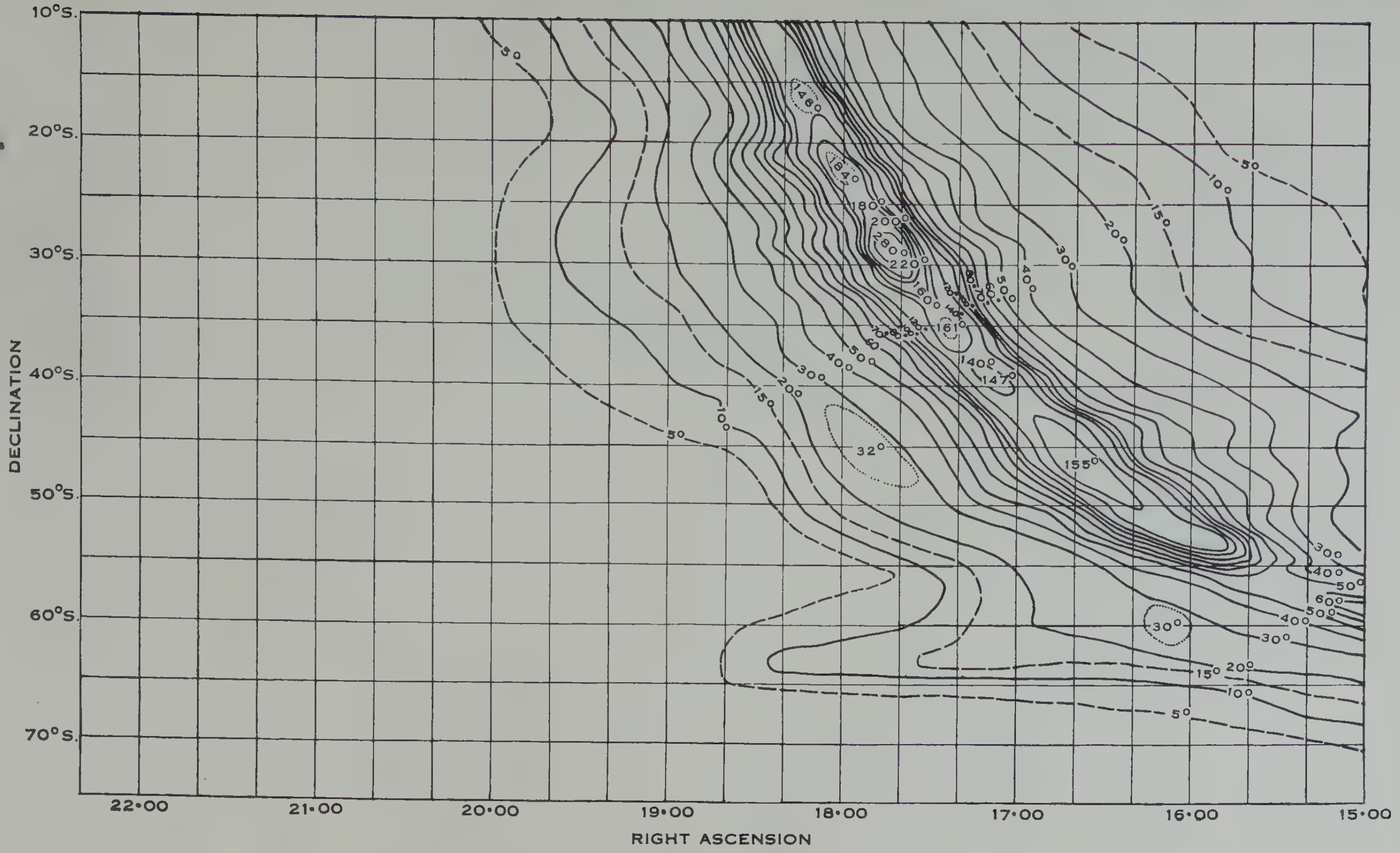


Fig. 1 (*c*).—Isophotes of cosmic radio emission. The southern plot R.A. 15^h to R.A. 22^h.

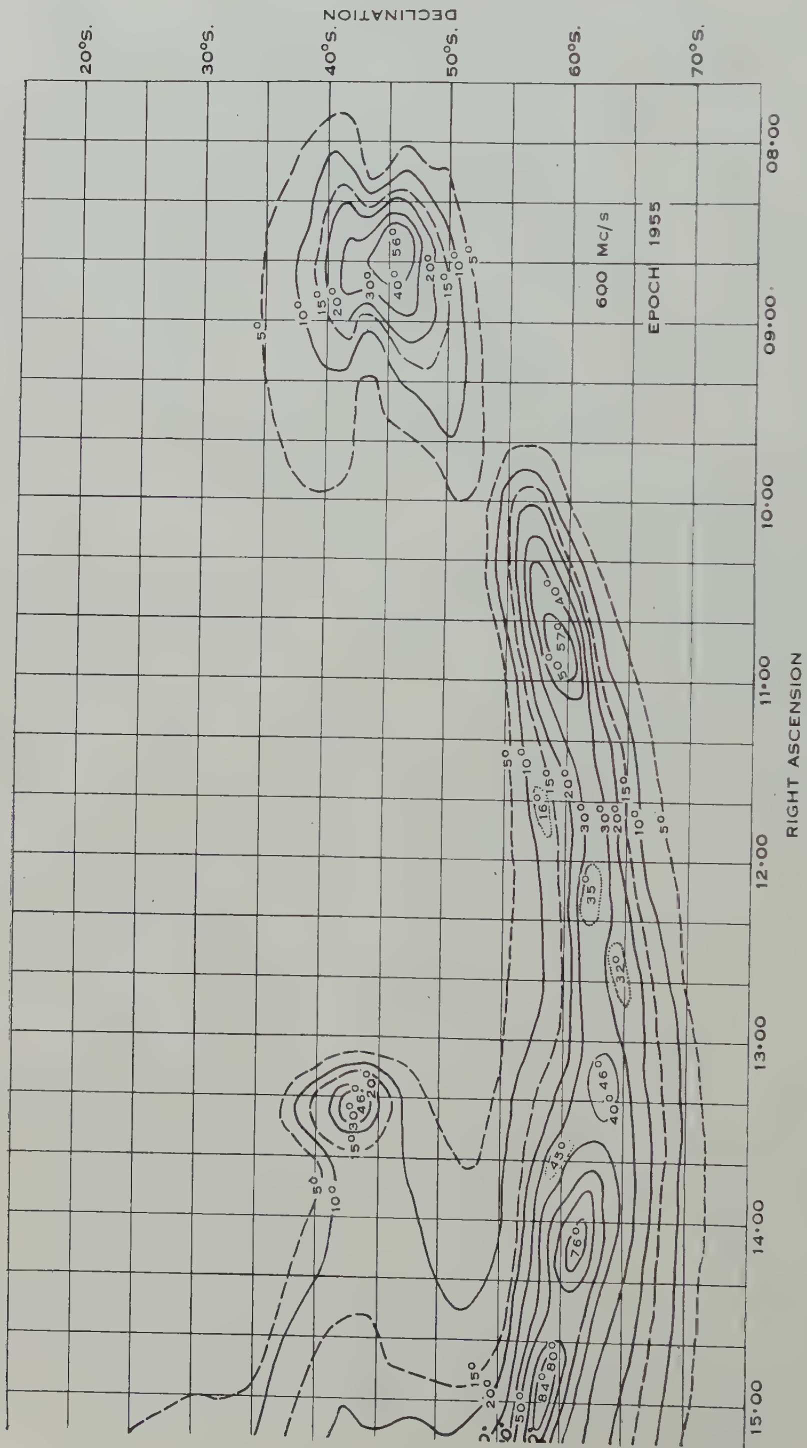


Fig. 1 (*d*).—Isophotes of cosmic radio emission. The southern plot R.A. 08h to R.A. 15h.

extends from about Dec. 2 °S. to 28 °N. while the present plots show a main region between Dec. 6 °S. and 17 °N. together with two or three discrete sources clustered at about 22 °N.

III. Accuracy of the Plot

Before discussing sources of error in the plot it is desirable to decide just what should be the quantity plotted. Broadly speaking, a well-designed aerial system of the type used in the present survey when used to transmit, radiates about two-thirds of the total power into the main beam. The remaining one-third is scattered over the whole of the rest of a sphere concentric with the aerial (into the so-called " side lobes "). Similarly, when receiving energy we may consider the equivalent aerial resistance as composed of two parts, the first, of two-thirds the total, being exposed to radiation from the main lobe and the second, of one-third, to radiation from the side lobes. Evidently, when the main lobe is directed at a region of brightness temperature T and the side lobes at regions of zero temperature, the measured aerial temperature is $2T/3$. What we call the " beam temperature " is then 3/2 times the measured aerial temperature, that is T.

This idealized case is never realized during a survey. Radiation from the ground and from some bright parts of the sky enters the side lobes at all times and the sum total of this radiation will vary as the aerial moves or as the Galaxy passes overhead. However, in the present survey, the aerial is never moved during an observation and only aerial temperature differences are measured. This completely cancels out the effects of ground radiation. It would be virtually impossible to integrate the varying contributions entering the side lobes from the Galaxy both directly and via ground reflection. To do so would require a detailed knowledge of the whole three-dimensional aerial pattern together with varying ground reflectivity, the whole to be combined with the as yet unknown plot of sky brightness. However, because of the wide scatter of the side lobes and the small angular extent of the bright parts of the sky, it is a reasonable approximation to assume that the energy entering the side lobes over a period required for the Galaxy to pass through the main beam is constant. The error involved is never likely to reach 10 per cent. for any of the brighter areas of the survey.

In the plot of Figure 1 changes in energy entering the aerial through side lobes are neglected and the measured aerial temperature is multiplied by a factor 100/65 (Piddington and Trent 1956) to give the plotted values of beam temperature.

The alternative method, of plotting equivalent aerial temperatures, has often been followed. In a personal communication Professor J. D. Kraus has kindly informed us that his intensity contours at 242 Mc/s (Kraus and Ko 1955) are neither aerial temperatures nor beam temperatures but approximately the mean value of the two. The plotted values are 1·26 times the aerial temperature, the beam temperature is 1·5 times the aerial temperature. This method might give more accurate values of brightness temperature in regions of low intensity. Summing up, it seems unimportant which method is adopted provided the method is clearly specified ; otherwise errors up to 50 per cent. are likely when making spectrum comparisons.

A second source of inaccuracy in the plot is the unknown base-level of radiation outside the 5 °K contour. In making the plot the zero level for each declination was that below which no decrease could be measured. Such a procedure is not justified at lower frequencies where variations in the minimum level are measurable. These measured variations at lower frequencies may be used to determine corrections at 600 Mc/s. The true zero level was estimated by combining the spectrum of emission from the coldest parts of the sky (near R.A. 09^h 50^m, Dec. +22° or R.A. 04^h 30^m, Dec. —40°) with a general plot of brightness at a lower frequency where brightness in these regions has been measured. The best available data are combined in Figure 2 to provide the required spectrum curve. If this curve is extrapolated to 600 Mc/s the brightness is about $4 \cdot 5 \times 10^{-22}$ W m^{-2} (c/s)$^{-1}$ sterad^{-1}, corresponding to a brightness temperature of 4 °K. This figure gives the base level at declinations where the aerial beam passes over the coldest parts of the Galaxy. At other declinations the base level will be slightly higher and may be estimated by comparing Figure 1 with a plot for a frequency at which absolute levels have been measured and assuming the distribution at the two frequencies broadly similar. The highest frequency at which such measurements are reasonably easy and accurate is

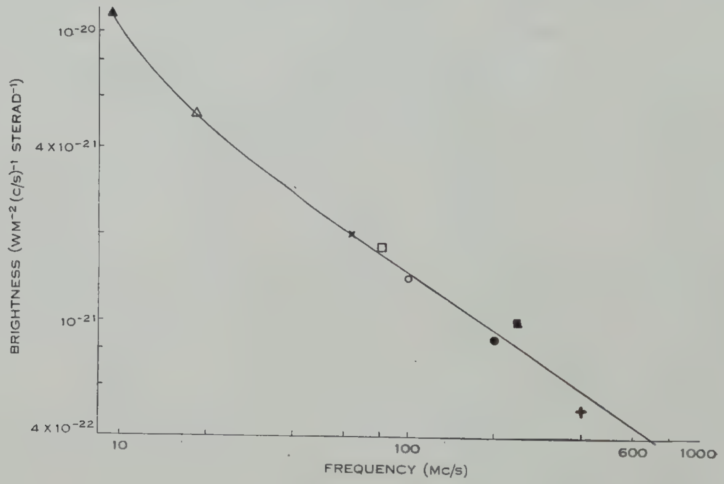


Fig. 2.—The observed cosmic radio spectrum from the coldest parts of the sky (near R.A. 09^h 50^m, Dec. +22° and R.A. 04^h 30^m, Dec. —37°) in units of brightness.

▲ Higgins and Shain (1954). △ Shain (1951). × Hey, Parsons, and Phillips (1948). □ Baldwin (1955). ○ Bolton and Westfold (1950). ● Piddington (1951), a theoretical estimate based on the survey of Allen and Gum (1950). ■ Kraus and Ko (1955). + McGee, Slee, and Stanley (1955).

about 100 Mc/s and so the isophotes of Bolton and Westfold (1950) are used.* The absolute levels were determined at points where our 5 °K isophotes intersected lines of constant declination of values 30 °N, 0°, 30 °S., and 60 °S. The levels, adjusted to 600 Mc/s by means of Figure 2, ranged from 9 to 13 °K, with an average value of 10·7 °K. Thus the base level ranged from 4 to 8 °K, with an average value of 5·7 °K.

To find absolute values of brightness temperature from Figure 1 a mean value of 5·7 °K may be added to the plotted value or the correction for the particular declination found as above ; it should lie between 4 and 8 °K.

IV. The Origin of the Radiation

There are at least two sources of galactic radio emission, one being ionized hydrogen emitting thermally (Piddington 1951). It is now fairly certain that the second source is relativistic electrons emitting by the synchrotron process (Alfvén and Herlofson 1950 ; Kiepenheuer 1950 ; Shklovskii 1953 ; and others). The interpretation of surveys is greatly complicated by the fact that the emitting hydrogen also acts as an absorber of both the thermal and non-thermal radiation, the effect being significant at 100 Mc/s and increasing rapidly with decreasing frequency.

Estimates of the temperature of the hot hydrogen (H II) regions agree at about 10^4 °K. The cold hydrogen (H I) regions may be at 100 °K or less but as they emit and absorb negligible amounts of 600 Mc/s energy their presence may be ignored. The maximum brightness temperatures found in the survey are less than 300 °K so that, if all the emission were from H II regions, the maximum optical depth, averaged over the beam, would be less than 0·03. Thus absorption is negligible except perhaps by H II clouds which subtend angles much less than the beamwidth ; these would obscure only small areas on the plot. It is reasonable to conclude that, within the limits of sensitivity of the equipment, all or nearly all galactic radiation is observed.

The proportions of thermal and non-thermal components in the observed radiation is more difficult to determine. From the early analysis (Piddington 1951) it was concluded that, at frequencies above a few hundred Mc/s, most of the radiation from regions near the galactic plane had a thermal origin. Objections to this conclusion were raised by Hanbury Brown (1953) on the grounds that the early observational data were not reliable and that the result is not consistent with visual evidence. He assumed, in developing a galactic model of radio sources, " that the majority of the galactic radiation observed towards the galactic centre at 480 Mc/s does not arise in ionized interstellar gas ". This assumption appears to be based on the fact that the observed components of Baade's Population I are rare in the nuclei of spiral nebulae similar to our Galaxy.

There is now no doubt that the earlier plots of galactic emission were in serious error. However, the conclusions regarding the origin of galactic radiation were based on general spectral trends, much of the data being obtained with aerial beams of more or less similar breadth. This tended to eliminate errors due to individual inaccurate plots so that the conclusions, while far from definite, should carry some weight. On the other hand Hanbury Brown's argument concerning galactic structure is based on optical evidence which is not yet adequate to determine the type of galaxy, so that inferences made from observations of " similar " galaxies are not justified.

---

* Although inaccurate near the galactic plane, these should be satisfactory in the regions of reasonably uniform brightness with which we are concerned.

More recently Mills (1955) has estimated the maximum proportion of the thermal component near the galactic centre. He divided the total observed radiation into a " spherical subsystem " and a " flattened subsystem " and found that, at 85 Mc/s, only 26 per cent. or less of the latter could be of thermal origin. Since the radiation from the spherical system is certainly non-thermal this would mean a proportion of thermal component in the total radiation of less than 20 per cent. If the non-thermal spectrum is similar to that shown in Figure 2 then the corresponding proportion of thermal radiation at say 200 Mc/s would be less than 32 per cent., in contrast with the earlier conclusion that most of the radiation was thermal. However, Mills's analysis seems open to question. To eliminate the contributions of the " spherical subsystem " and of the discrete source in this direction he divided the observed intensities at 85 Mc/s, 400 Mc/s, and 1210 Mc/s by factors of about 1·3, 2·6, and 1·0 respectively. Such estimated factors are subject to considerable uncertainty which is reflected in the derived ratio of the components. Thus if the observed intensities are compared directly,* then the possible proportion of thermal radiation at 85 Mc/s could be about 50 per cent.

This raises the question of the division of the non-thermal galactic radiation into the two subsystems mentioned above. As evidence Mills (1955) states that the galactic sections of his Figure 6, with the sharp peaks (the " flattened subsystem ") removed, all have maxima at the Right Ascension of the galactic centre. This conclusion depends on the amount of peak which is shaved off. An alternative interpretation which appears to fit the data equally well is that the non-thermal emission constitutes a single spheroidal system of variable emissivity per unit volume, the latter increasing towards the galactic plane where the thermal radiation is also concentrated. This interpretation lies between those of Shklovskii (1953), who perhaps overestimated the part played by the thermal component, and Mills, who may have underestimated it.

At the galactic crossing at 25½ °S. (a region of high intensity yet reasonably free of discrete sources) a comparison was made of the brightness at frequencies of 242 Mc/s (Kraus and Ko 1955), 400 Mc/s (McGee, Slee, and Stanley 1955), and 600 Mc/s (the present survey). To obtain beam temperatures the 242 Mc/s data were multiplied by 1·19 (see Section III) and the 400 Mc/s data by 1·5

* The 1210 Mc/s data with which one of the authors was associated (Piddington and Minnett 1951) are *aerial* temperatures. The conversion factor to beam temperatures was never determined but is likely to be much higher than 1·5 because the feed dipole was located away from the focal point and among several other dipoles. Furthermore, the reflector was subject to sag and distortion. It is undesirable, therefore, to use these data in determining spectra.

(a factor kindly provided by Mr. R. X. McGee, personal communication). The intensities in power units fall within ±4 per cent., the 242 Mc/s value being the lowest and the 600 Mc/s value the highest. This result provides considerable evidence in favour of a preponderance of thermal radiation. Some measurements at 900 Mc/s (Denisse, Leroux, and Steinberg 1955) do not appear to be consistent with such a conclusion, however, unless the aerial beam used was much wider than at the lower frequencies. The 900 Mc/s intensity falls well below those at the other frequencies.

Optical evidence (the absence of H II clouds) shows that emission from regions well away from the galactic plane must be non-thermal.

V. The Radio Galactic Equator and Position of the Sun

The combination of the narrow beam used and the wide extent of the present survey may make it useful in redetermining the position of the galactic equator and estimating the distance of the Sun above the galactic plane. If the conclusions of the last section are correct then the "galaxy" in question is largely the H II distribution.

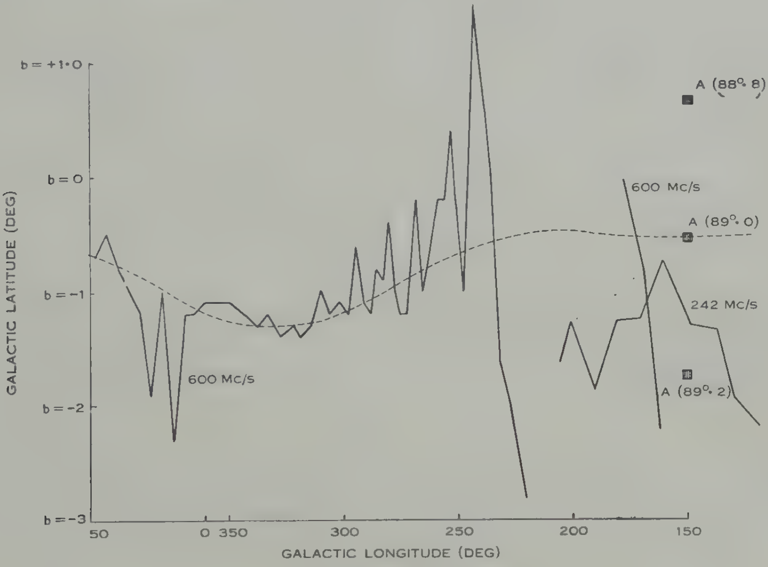


FIG. 3

Fig. 3.—The position of the ridges of maximum brightness at 600 and 242 Mc/s and of a hypothetical galaxy (galactic coordinates, Lund pole).

In Figure 3 the position of the ridge of maximum intensity of radio emission is plotted in galactic coordinates; the Lund tables were used and correction made for changes in the celestial coordinates since epoch 1900. The plot covers longitudes between 50°, through zero, to 220° and is shown by the full line marked 600 Mc/s. The other lines are discussed below. It is evident that only within about ±40° of the galactic centre (near $l \sim 330°$) is the curve smooth.

This is because radiation from this region is the integrated emission from numerous sources lying at all distances up to perhaps 20,000 parsecs. Beyond this sector the plot is increasingly irregular since the radiation originates in fewer and closer sources. The statistical value of the plot in the region $l \sim 330°$ is very much greater than elsewhere. The maximum displacement between the Lund equator and the 600 Mc/s radio equator lies in this sector and so may be determined with considerable accuracy. However, the galactic longitude at which this maximum occurs cannot be accurately determined. The displacement is 1°·3 with an error probably not more than 0°·1; this maximum occurs near $l=330°$.

The equatorial displacement of 1°·3 is due to the combination of two factors: the radio pole is displaced from the Lund pole and the Sun is displaced from the radio equatorial plane. It is desirable to estimate the approximate relative importance of the two factors; this is done on the assumption that the Galaxy is circularly symmetrical and of diameter 19,000 parsecs with the Sun 1600 parsecs from the outer edge.

A configuration corresponding to the radio pole lying at $b=89°\cdot 0$ (and $l \sim 330°$) is shown by the dashed curve of Figure 3. The corresponding position of the Sun is 42 parsecs from the equatorial plane. Corresponding curves with the pole at 88°·8 and 89°·2 would coincide with the dashed curve near $l=330°$ but diverge to pass through the squares marked A (88°·8) and A (89°·2). The dashed curve appears to give the best fit.

Also shown in Figure 3 is a plot of the radio equator near the galactic anticentre obtained at 242 Mc/s by Ko and Kraus (1955). These data favour a displacement of the galactic pole to about 89°·1 with a solar distance of 56 parsecs.

Either of the above polar positions agrees well with that of van Tulder (1942); from a wide array of optical evidence he set the pole at $l=328°$, $b=89°\cdot 0$. His solar displacement was much smaller, however, at 13·5 parsecs.

VI. ACKNOWLEDGMENTS

The authors would like to thank Mr. F. J. Kerr, Mr. B. Y. Mills, and Mr. C. A. Shain for helpful suggestions regarding the manuscript; and Mr. F. J. Kerr for provision of conversion charts used to change to galactic coordinates.

VII. REFERENCES

Alfvén, H., and Herlofson, N. (1950).—*Phys. Rev.* **78**: 616.
Allen, C. W., and Gum, C. S. (1950).—*Aust. J. Sci. Res.* A **3**: 224.
Baldwin, J. E. (1955).—*Observatory* **75**: 229.
Bolton, J. G., and Westfold, K. C. (1950).—*Aust. J. Sci. Res.* A **3**: 19.
Bracewell, R. N., and Roberts, J. A. (1954).—*Aust. J. Phys.* **7**: 615.
Denisse, J. F., Leroux, E., and Steinberg, J. L. (1955).—*C.R. Acad. Sci., Paris* **240**: 278.
Hanbury Brown, R. (1953).—*Phil. Mag.* **44**: 939.
Hey, J. S., Parsons, S. J., and Phillips, J. W. (1948).—*Proc. Roy. Soc.* A **192**: 425.
Higgins, C. S., and Shain, C. A. (1954).—*Aust. J. Phys.* **7**: 460.
Kiepenheuer, K. O. (1950).—*Phys. Rev.* **79**: 738.

Ko, H. C., and Kraus, J. D. (1955).—Rep. Ohio Univ. Radio Obs. No. 4.
Kraus, J. D., and Ko, H. C. (1955).—*Nature* **175** : 159.
McGee, R. X., Slee, O. B., and Stanley, G. J. (1955).—*Aust. J. Phys.* **8** : 347.
Mills, B. Y. (1955).—*Aust. J. Phys.* **8** : 368.
Piddington, J. H. (1951).—*Mon. Not. R. Astr. Soc.* **111** : 45.
Piddington, J. H., and Minnett, H. C. (1951).—*Aust. J. Sci. Res.* A **4** : 459.
Piddington, J. H., and Trent, G. H. (1956).—*Aust. J. Phys.* **9** : 74.
Reber, G. (1948).—*Proc. Inst. Radio Engrs., N.Y.* **36** : 1215.
Shain, C. A. (1951).—*Aust. J. Sci. Res.* A **4** : 258.
Skhlovskii, I. S. (1953).—*Astr. J., Moscow* **30** : 15.
van Tulder, F. F. M. (1942).—Bull. Astr. Insts. Netherlds. No. 353 : 1.

AN INVESTIGATION OF THE RADIO SOURCE 06N2A IN GEMINI

By H. Rishbeth*

[*Manuscript received June* 18, 1956]

Summary

A lunar occultation of the radio source 06N2A in Gemini was observed in the course of an investigation at a wavelength of 3·5 m, for which the " Mills Cross " aerial was employed. During the occultation the intensity of the source was reduced by over one-fifth, although less than one-tenth of the associated galactic nebulosity IC 443 was obscured. The radio source appears to resemble the nebula, and the relation between them is discussed with the aid of simple geometrical models of the source.

A comparison of the intensity of the source with previously published values shows that the emission at metre wavelengths is mainly non-thermal.

I. Introduction

During 1955 and early 1956 there occurred a series of lunar occultations of the radio source 06N2A in Gemini, and on October 8, 1955 the source culminated at Sydney during the maximum phase of one of these events. It was decided to carry out an investigation of the source, including observations of the occultation, using the 1500-ft " Mills Cross " which operates at a wavelength of 3·5 m.

It was pointed out by Link and Neužil (1954) that lunar occultations provide high resolution for studying the brightness distributions of radio sources. The method is particularly valuable when applied to sources such as 06N2A which, though of appreciable angular size, are too small to be investigated in detail at metre wavelengths by ordinary observations with existing radio telescopes.

The radio source 06N2A in Gemini has been identified with the galactic nebulosity IC 443 (Baldwin and Dewhirst 1954). A photograph of this object, taken in red light with the 48-in. Schmidt camera at Palomar as part of the National Geographic Society-Palomar Observatory Sky Survey, and kindly presented by Dr. R. Minkowski, shows that the nebula is roughly circular in shape and about 48 min of arc in diameter (Fig. 1). The optically bright regions show a filamentary structure : the brightest forms an arch at the sharply defined north-following boundary of the nebula, and other bright portions lie toward the south-preceding edge. The central regions of the nebula are partially obscured.

The first observation of a lunar occultation to be reported was that of Elsmore and Whitfield (1955), who observed an earlier occultation of 06N2A at wavelengths of 3·7 and 7·9 m.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Chippendale, N.S.W.

The results of the present investigation are described in Section II. Their interpretation involves the aerial characteristics and the radiation from the galactic background and the Moon, and it is necessary to discuss these influences before comparing the optical and radio objects, as is done in subsequent sections.

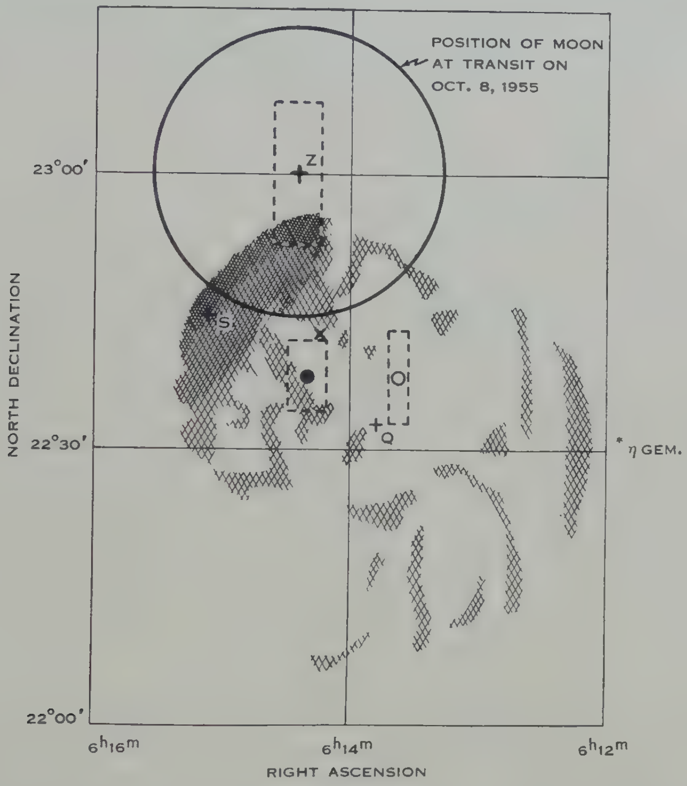


Fig. 1.—Diagram of the nebula IC 443 and the radio source 06N2A. The position of the centroid of the aerial response to the source is denoted by a solid circle; the peak by a cross (×); and the position obtained by Baldwin and Dewhirst by an open circle (1950·0 coordinates). The other positions shown are as follows: *Z*, centre of occulted radio brightness; *Q*, centre of IC 443, and of extended source *Q* in model of Section IV (*b*); *S*, centre of source *S* in model of Section IV (*b*); *, star η Geminorum. The areas enclosed by broken lines indicate the probable errors.

II. Observations

The instrument employed was the Mills Cross aerial at Fleurs, near St. Mary's, N.S.W., of which a detailed description is being prepared by Mills and others. The aerial is in the form of a cross whose arms are 1500 ft long, and has a pencil-beam response which is circular and of half-power width 49′ when pointed to the zenith, but which becomes elongated when the beam is swung to lower elevations owing to the foreshortening of the array. The principle of operation of the instrument has been described by Mills and Little (1953).

The aerial receives horizontally polarized radiation, and during a switching cycle of 1 min duration the beam is pointed successively in five adjacent directions in the meridian. Throughout the observations the spacing between these five positions was about 37 min of arc, which was roughly two-fifths of the beam-width in declination. On four days, including the day of the occultation, the central beam position was directed to Dec. +23° 05′; on two other days, to Dec. +21° 53′; and on one further day, to Dec. +24° 05′. The aerial beam cannot easily be swung in azimuth, and so it was not possible to follow the progress of the occultation. The intense radio source 05N2A in Taurus (M1, the Crab Nebula) passed through the aerial beam about three-quarters of an hour before 06N2A, and provided a check on the directional characteristics of the aerial.

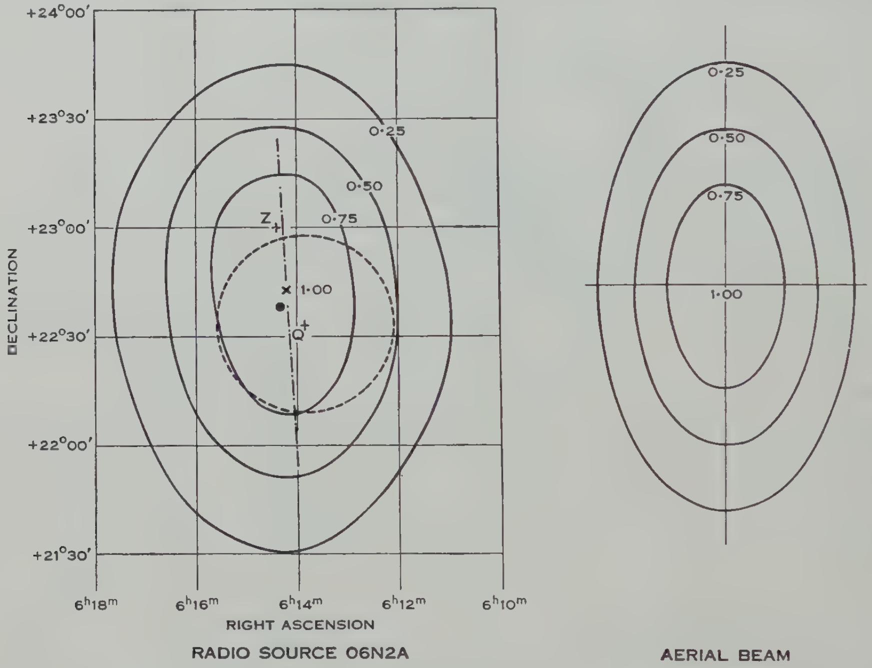


Fig. 2.—Contours of equivalent aerial temperature relative to the peak value for source 06N2A (1950·0 coordinates). As in Figure 1, the centroid of the aerial response is denoted by a solid circle and the peak by a cross (×). *Q* is the centre of IC 443 and *Z* the centre of occulted radio brightness. The equivalent aerial temperature at the peak is 4700 °K. The intensity contours of the aerial beam are also shown.

– – – – Approximate boundary of IC 443.

·—·—·— Major axis of observed aerial response.

The position of the Moon at transit on the day of the occultation was computed from the data of Link and Neužil and information supplied by the staff of " The Nautical Almanac ". The relation of this position to the nebula IC 443 and the radio source is shown in Figure 1. Only about one-tenth of the area of IC 443 was covered by the Moon; nevertheless, the radio intensity was reduced by about one-fifth. The centre *Z* of occulted radio brightness was found by comparing the record obtained during the occultation with those obtained on the control days; it lies further north than expected, but the uncertainty of the measurement allows of a more plausible position within the nebula. The dimensions of the occulted region were small compared to the aerial beam.

The principal results of the investigation at 3·5 m, with their estimated probable errors, are presented numerically in Table 1, and the observed contours of radio brightness of the source 06N2A in the absence of the Moon are shown in Figure 2.

TABLE 1

PRINCIPAL RESULTS AT 3·5 M

Property	06N2A Gemini	05N2A Taurus
Observed radio position (1950·0) Peak	$6^{h}\,14^{m}\cdot2\pm0^{m}\cdot15$, $+22^{\circ}\,43'\pm04'$	$5^{h}\,31^{m}\cdot4\pm0^{m}\cdot1$, $+21^{\circ}\,59'\pm03'$
Centroid	$6^{h}\,14^{m}\cdot3\pm0^{m}\cdot15$, $+22^{\circ}\,38'\pm04'$	$5^{h}\,31^{m}\cdot5\pm0^{m}\cdot1$, $+21^{\circ}\,59'\pm03'$
Position of optical centre (1950·0)	$6^{h}\,13^{m}\cdot8$, $+22^{\circ}\,33'$	$5^{h}\,31^{m}\cdot50$, $+21^{\circ}\,59'\cdot3$
Half-power width of aerial response to source (east-west × north-south)	$66'\times99'$	$49'\times86'$
Computed half-power width of aerial beam	$49'\times83'$	$49'\times82'$
Diameters of optical object	$48'\times48'$	$\ll$beamwidth
Flux density (W m^{-2} (c/s)$^{-1}$)	$(66\pm15)\times10^{-25}$	$(230\pm35)\times10^{-25}$
Brightness temperature of equivalent disk, 48′ in diameter (°K)	14500	—
Proportion of radio intensity occulted by Moon	$0\cdot22\pm0\cdot05$	—
Flux density of occulted portion (W m^{-2} (c/s)$^{-1}$)	$(15\pm4)\times10^{-25}$	—
Centre of occulted portion (1950·0)	$6^{h}\,14^{m}\cdot4\pm0^{m}\cdot2$, $+23^{\circ}\,00'\pm08'$	—
Topocentric position of Moon's centre at transit (1950·0)	$6^{h}\,14^{m}\cdot4$, $+23^{\circ}\,01'$	—

The probable errors quoted are principally due to uncertainty in the performance and absolute calibration of the aerial, and the scatter of the daily observations do not contribute greatly to these errors. For instance, all the measured positions of the peak of 06N2A lie within a range of $0^{m}\cdot3$ in Right Ascension; and the total scatter of the daily values of flux density of this source was but 10 per cent. of the mean value. Ratios of flux density are thus more reliable than their absolute values.

The correction which was applied for refraction amounted to 3′ in declination, to which the troposphere and ionosphere contributed 2′ and 1′ respectively. Refraction in Right Ascension due to the ionospheric wedge effect was neglected, for the greatest deviation observed at 3·7 m by Smith (1952) is of order $0^{\mathrm{m}}\cdot 05$. The angular size of the diffraction pattern around the lunar disk is of order 1′; this is too small to influence significantly the reduction of intensity observed during the occultation. Solar interference was absent as the sources culminated before dawn, but ionospheric irregularities caused severe scintillations on one of the control days.

Observations of the source at a wavelength of 50 cm were made by G. Trent of the Radiophysics Laboratory, using a transit aerial at Potts Hill, Sydney. No reduction of intensity on the day of the occultation was observed, but the sensitivity of the system was such that a decrease of 20 per cent. would barely be detectable.

The occultation of 06N2A on January 25, 1956 was also visible at Sydney.* Unfortunately, a violent thunderstorm coincided with the event, and rendered useless the observations made by A. W. L. Carter of the Radiophysics Laboratory and the author.

For comparison, a summary of previous observations of 06N2A are given in Table 2.

TABLE 2

SUMMARY OF PREVIOUS OBSERVATIONS OF 06N2A

Reference	Wave-length (m)	Fre-quency (Mc/s)	R.A.	Dec. (north)	Flux Density (10^{-25} W m^{-2} (c/s)$^{-1}$)
Haddock, Mayer, and Sloanaker (1954) . .	0·094	3200	$6^{\mathrm{h}}\ 14^{\mathrm{m}}$	22° 30′	5·5
Hagen, Lilley, and McClain (1955) . .	0·21	1420	$6^{\mathrm{h}}\ 13^{\mathrm{m}}$	22° 40′	16
Piddington and Trent (personal communication)	0·50	600	$6^{\mathrm{h}}\ 13^{\mathrm{m}}$	23°	22
Ko and Kraus (1955)	1·24	242	$6^{\mathrm{h}}\ 15^{\mathrm{m}}\pm 1^{\mathrm{m}}$	$23^{\circ}\pm 1^{\circ}$	16
Rishbeth (see Table 1)	3·5	85·5	$6^{\mathrm{h}}\ 14^{\mathrm{m}}\cdot 3\pm 0^{\mathrm{m}}\cdot 15$	$22^{\circ}\ 38'\pm 4'$	66
Baldwin and Dewhirst (1954) . .	3·7	81·5	$6^{\mathrm{h}}\ 13^{\mathrm{m}}\ 37^{\mathrm{s}}\pm 4^{\mathrm{s}}$	$22^{\circ}\ 38'\pm 5'$	42
Baldwin and Dewhirst (1954) . .	7·9	38			65

III. FACTORS INFLUENCING THE OBSERVATIONS

Before the observed brightness distribution of the source is discussed in detail, it is necessary to examine certain factors which affect the observations. The performance of the aerial is of prime importance, because of the similarity of angular size between the aerial beam and the source. The results are also influenced by the galactic background in the neighbourhood of the source, and by the brightness temperature of the Moon itself.

---

* The quantity *H* (Greenwich Hour Angle of the source at the instant of its conjunction with the Moon) is given incorrectly by Link and Neužil for this occultation and two others of 06N2A in 1956.

(*a*) *The Performance of the Aerial*

The source 06N2A transits at a zenith angle of 57°, 12° beyond the normal operating range of the aerial. Under these conditions no detailed analysis of the behaviour of the aerial has been made, but it is known that the sensitivity is somewhat reduced and the beam elongated in declination.* Taurus-A (05N2A) may be considered as a point source when viewed by the aerial, and the observed beamwidth at half-power in declination is 86′, in fair agreement with the calculated value of 82′ (see Table 1). The east-west beamwidth agrees with the theoretical width of 49′, which should not vary with zenith angle. The adopted beamwidths at the declination of 06N2A are 49 and 87′.

The position of the centroid of the aerial response to 05N2A lies at the centre of the nebula M1, and so the collimation error of the aerial must be small notwithstanding the large zenith angle. For this reason the position quoted in Table 1 for the centroid of 06N2A is considered reliable, even though the position given by Baldwin and Dewhirst differs from it by $0^{m}\cdot 7$ in Right Ascension. There is, however, a displacement of $1\frac{1}{2}\pm\frac{1}{2}$ min of arc in azimuth between the peak and centroid of the response to 05N2A. As the latter is effectively a point source, this displacement indicates that the aerial beam is asymmetrical; possibly as the result of ground reflections which might become serious at such large zenith angles. Consequently, the displacement in Right Ascension between the peak and centroid of the Gemini source may not be real.

(*b*) *The Effect of the Galactic Background*

The source 06N2A is situated in a region of strong emission near the galactic anticentre. The records of the source have been analysed in detail over the range from $6^{h}\,00^{m}$ to $6^{h}\,30^{m}$ in Right Ascension, and the regions surrounding 06N2A give an almost uniform response, from which the deflection due to the discrete source is easily separated. Thus the galactic structure in the neighbourhood of the source is unlikely to affect the accuracy of the results. Confusion with another discrete source could significantly affect the observations if such a source exceeded about 5×10^{-25} W m^{-2} (c/s)$^{-1}$ in intensity and lay within an area of roughly 4 square degrees centred on the position of 06N2A. Results of earlier surveys enable the chance of this occurrence to be estimated as 1 : 60. Such confusion might possibly cause the discrepancy between the position obtained and that of Baldwin and Dewhirst, as their observations with an interferometer possessing a wide primary beam are more prone to confusion with another nearby source than the author's observations with a pencil-beam aerial.

(*c*) *The Effect on the Occultation of the Apparent Brightness Temperature of the Moon*

The Moon may be considered as a uniform disk, of which the apparent brightness temperature is negative. This arises from the obscuration of the galactic background, whose brightness temperature was found to be about

---

* The Mills Cross has recently been recalibrated by A. G. Little of the Radiophysics Laboratory. His measurement of the flux density of 05N2A is quoted in Table 1, and the other flux densities are based upon it.

1500 °K. This disk temperature is made less negative by two terms: firstly, that due to the reflection of galactic radiation by the Moon; and, secondly, the temperature of the lunar surface, which, according to the $1 \cdot 25$ cm observations of Piddington and Minnett (1949), is 270 °K at the Moon's last quarter. This lunar radiation is thermal, and so the effective brightness temperature at $3 \cdot 5$ m is likely to be similar. The magnitude of the former effect can be estimated as follows. The Moon is assumed to be a diffuse reflector of directivity 5 relative to an isotropic radiator and of reflection coefficient $0 \cdot 15$ (Kerr and Shain 1951). Then the effective brightness temperature of the reflected radiation is roughly $0 \cdot 15$ of the average brightness temperature of an area of sky opposite the Moon and covering one-fifth of the celestial sphere.

No complete survey of this region is available at a wavelength of $3 \cdot 5$ m, but the required temperature is estimated to be 1500 °K at $3 \cdot 0$ m using the contours of Bolton and Westfold (1950). At $3 \cdot 5$ m the value is likely to be higher by a factor of about $1 \cdot 5$. Reflected solar radiation is unlikely to contribute greatly to this as no solar activity was present on the day of the occultation. Thus the estimated brightness temperature of reflected radiation from the Moon is $0 \cdot 15 \times 1 \cdot 5 \times 1500$ °K or 340 °K, giving a final value of 900 °K for the negative apparent disk temperature of the Moon, corresponding to a negative flux density not exceeding $1 \cdot 5 \times 10^{-25}$ W m^{-2} (c/s)$^{-1}$. The presence of an extraneous discrete source within the occulted area would not greatly affect the result unless its intensity exceeded 2×10^{-25} W m^{-2} (c/s)$^{-1}$; considerations similar to those employed in Section III (*b*) give the chance of this circumstance as about 1 : 250. It is thus almost certain that the total flux density of occulted background is small compared to the reduction of 15×10^{-25} W m^{-2} (c/s)$^{-1}$ observed, and the balance is attributed to the obscuration of the discrete source 06N2A.

IV. Discussion

(*a*) *Contours of Equivalent Aerial Temperature*

A contour diagram of equivalent aerial temperature of the source 06N2A is shown in Figure 2. The portions outside the quarter-power contour are indefinite and have been omitted from the determination of the position of the centroid, though they have been included in the value obtained for the flux density.

The asymmetry of the aerial beam, discussed in Section III (*a*), may be sufficient to account for the observed skewness of the east-west section of the aerial diagram (Fig. 3). The shape of the aerial beam in the north-south direction is not accurately known, but the close spacing of the contours on the northern side shows that the source has a northern boundary sharp compared to the aerial beam.

When the diagram is compared with the position of IC 443, the centroid is found to lie well toward the following side of the nebula, and slightly to the north of the optical centre. This indicates a concentration of radiation in the north-following quarter, in the neighbourhood of the brightest optical emission. Moreover, the observations during the occultation show that over one-fifth of the radiation originates in the small part of the bright north-following arch, occupying less than one-tenth of the total area of IC 443, that was covered by the Moon. This arch was not obscured during the occultation observed by Elsmore and Whitfield, and these authors postulated a concentration of radiation in its vicinity on the ground that the observed reduction in intensity of the source was less than that expected from the circularly symmetrical model of Baldwin and Dewhirst.

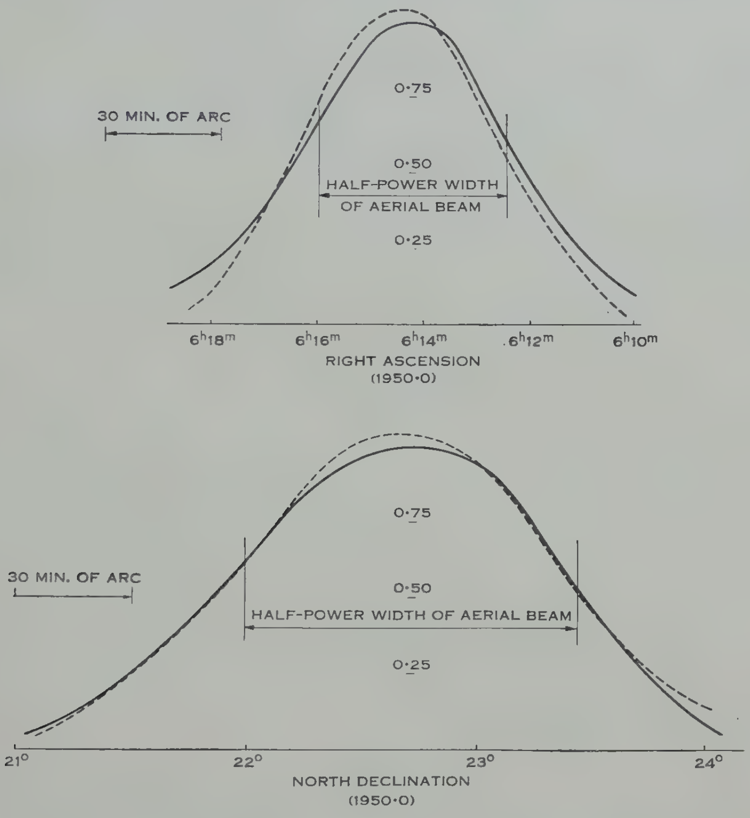


Fig. 3.—Sections through the peak of the contours of Figure 2 for source 06N2A, with those derived from the model of Section IV (*b*). (1950·0 coordinates.)

——— Observed section.

- - - - - Section derived from model of uniform disk *Q* and small source *S*.

The major axis of the contour diagram is slightly inclined to the meridian : the peak of an east-west section through the northern extremity of the half-power contour is about $0^{m}\cdot 3$ later in Right Ascension than the peak of the corresponding section at the southern extremity.

The extent of the contours shows that the source 06N2A is similar in size to the nebula IC 443, whose diameter is 48 min of arc. Baldwin and Dewhirst quoted the half-power width of the source as 20 min of arc; their distribution was, however, derived from measurements of the amplitude but not the phase of the Fourier components of the east-west strip brightness, and they assumed that the source possesses circular symmetry. A source of half-power width 20 min of arc would give a much narrower response to the beam of the Mills Cross than that obtained by the author.

(*b*) *Elementary Models of the Source*

A simple model of the radio brightness distribution across the source has been considered. It combines an extended source Q possessing circular symmetry, centred at the geometrical centre of IC 443, with a source S which is sufficiently localized to resemble a point source when scanned by the aerial.

The position of S and the intensities of S and Q were chosen to give the measured value of total flux density and the observed position of the centroid. The extended source Q was represented by a uniformly emitting disk whose diameter, 48′, is roughly that of IC 443; and the optimum ratio of the intensities of S and Q was found to be 3 : 4.

This model, when smoothed by a Gaussian aerial beam, reproduced moderately well some of the features of the observed distribution (Fig. 3). It failed near the peak, where the response might be expected to be sensitive both to the structure of the source, which the model ignores, and to the aerial beam asymmetry; it also failed to give a sufficiently wide response in the east-west direction. This suggested that the radiation might be more intense towards the boundaries of the nebula, and so another simple model was tried in which the source Q was taken as a uniform ring around the boundary of IC 443, augmented as before by a small source S. This model gave a slightly broader response in the east-west direction and a more accurate value of equivalent aerial temperature at the peak, but was still unsatisfactory in shape. Some improvement to both models could be made by enhancing the radiation originating in the south-preceding quarter.

The models can qualitatively account for the inclination of the major axis of the contour diagram to the hour circle through the peak (Section IV (*a*)), but they are not consistent with the observed displacement in Right Ascension between the peak and the centroid. It would seem difficult to find a model which does reproduce this displacement and also the other features of the response, and which does not require emission from outside the boundary of IC 443. However, as suggested in Section III (*a*), the separation of the peak and centroid may be instrumental in origin.

If the results given by the models are meaningful, the source S would correspond to the bright arch on the north-following boundary of IC 443. The dimensions of this arch are much smaller than the aerial beamwidths, and the response to a source extending along it would not differ greatly from that of a point source. The amount of radiation cut off by the Moon is consistent with the identification of S with the bright arch. It seems possible that the bright filamentary structures in the south-preceding quarter may also emit strong radiation.

(c) *Nature of the Radio Emission from IC 443*

In Table 2 are quoted several previous measurements of the flux density of 06N2A. If the source were thermal and optically thin, its flux density would not vary with wavelength. It is difficult to measure accurately the flux density of an extended source and it is possible that thermal radiation is dominant at short wavelengths. At metre wavelengths, however, the flux density rises sharply with increasing wavelength, and so a strong non-thermal component is present.

The source seems to belong to the class of faint filamentary diffuse objects with strong non-thermal radio emission. These objects sometimes have sharp outer boundaries, and possess violently moving filaments in which it is presumed that radio waves are generated by non-thermal processes. This type of source seems distinct from sources such as the Orion Nebula (M42) and other galactic H II regions which are optically bright masses of ionized gas, in which the internal velocities are 10 or even 100 times less than the first kind. There is at present no evidence that the radio emission from such objects is other than thermal (Mills, Little, and Sheridan 1956).

V. Conclusions

An analysis of the data has indicated that at 3·5 m the source 06N2A must be similar in size to the nebula IC 443, and that there is an apparent resemblance between the optical and radio objects. It has established the presence of strong radio emission in the neighbourhood of the bright arch on the north-following edge, and suggested that the filaments in the south-preceding quarter may also be strong radio sources. It has suggested the occurrence of " limb-brightening " at a wavelength of 3·5 m, such as might be expected were the nebula an expanding shell of gas—perhaps the remnant of a nova or supernova outburst (Shklovskii 1954). The resemblance between the radio and optical objects may be less real than suggested, as the nebula IC 443 is heavily obscured in places, notably near the centre, and the true distribution of optical brightness may be very different from that observed.

It has been assumed throughout that the radio emission is confined within the visible extent of IC 443. This seems to be a reasonable requirement as the nebula has a well-defined boundary. Filamentary structures exist well outside the nebula on the north-following side, but they are of a type fairly common in the Galaxy and there seems to be no reason for supposing them to be associated with the radio emission from IC 443.

Although the measured flux densities of the source at shorter radio wavelengths may be compatible with the presence of an optically thin thermal source, there is no doubt that at longer wavelengths the flux density rises rapidly and the emission is mainly non-thermal. The source appears to resemble the filamentary non-thermal objects such as Cassiopeia-A rather than the hot ionized gas clouds such as the galactic H II regions.

VI. Acknowledgments

The author is indebted to his colleagues for their assistance, and is especially grateful to Mr. B. Y. Mills and Mr. C. A. Shain for help and criticism. He would like to thank Dr. R. Minkowski for a helpful discussion, and Dr. J. H. Piddington for permission to quote his experimental results. The staff of " The Nautical Almanac " kindly supplied details of occultations visible at Sydney.

The work described in this paper was carried out during the tenure of a Research Studentship awarded jointly by the Commonwealth Scientific and Industrial Research Organization in Australia and by the Department of Scientific and Industrial Research in Great Britain.

VII. References

Baldwin, J. E., and Dewhirst, D. W. (1954).—*Nature* **173** : 164.
Bolton, J. G., and Westfold, K. C. (1950).—*Aust. J. Sci. Res.* A **3** : 19.
Elsmore, B., and Whitfield, G. R. (1955).—*Nature* **176** : 457.
Haddock, F. T., Mayer, C. H., and Sloanaker, R. M. (1954).—*Nature* **174** : 176.
Hagen, J. P., Lilley, A. E., and McClain, E. F. (1955).—*Astrophys. J.* **122** : 361.
Kerr, F. J., and Shain, C. A. (1951).—*Proc. Inst. Radio Engrs., N.Y.* **39** : 230.
Ko, H. C., and Kraus, J. D. (1955).—Rep. Ohio Univ. Radio Obs. No. 4.
Link, F., and Neužil, L. (1954).—*Bull. Astr. Insts. Csl.* **5** : 112.
Mills, B. Y., and Little, A. G. (1953).—*Aust. J. Phys.* **6** : 272.
Mills, B. Y., Little, A. G., and Sheridan, K. V. (1956).—*Aust. J. Phys.* **9** : 218.
Piddington, J. H., and Minnett, H. C. (1949).—*Aust. J. Sci. Res.* A **2** : 63.
Shklovskii, I. S. (1954).—*C.R. U.R.S.S. Acad. Sci.* **97** : 53.
Smith, F. G. (1952).—*J. Atmos. Terr. Phys.* **2** : 350.

THE DIURNAL VARIATION OF THE BAROMETER COEFFICIENT FOR COSMIC RAYS AT HOBART

By R. M. Jacklyn*

[*Manuscript received January* 9, 1956]

Summary

The records from a vertical counter telescope measuring the hard component of cosmic radiation at sea-level have disclosed significant diurnal variations of the barometer coefficient at Hobart, Tas. The amplitude of the variation is about 5 per cent., and there are secular changes of the same order during the mean day.

It is shown that it is meaningless to correct the diurnal variations of intensity to conditions of standard surface pressure using a barometer coefficient averaged over the mean day if the standard pressure chosen is very different from the observed mean value. Moreover, there is evidence from the Hobart results for 1954 that the observed diurnal maximum of the corrected intensities is associated with a low barometer coefficient and the minimum with a high coefficient, suggesting that the form of the corrected intensity variations may be determined largely by variations occurring within the atmosphere. Possible reasons for the variation of the barometer coefficient are discussed and an experiment is suggested which might decide what is the main cause of these variations, a changing primary spectrum or changing properties of the atmosphere.

I. Introduction

The main purpose of the study of the diurnal variations of cosmic ray intensity at the Earth's surface is to make deductions concerning the variations in solar time of primary cosmic rays at the top of the atmosphere. However, before this can be done it is necessary to take into account the effect of diurnal variations of the atmosphere on the cosmic ray intensity.

The effect produced by the atmosphere may come about in various ways. Consider first a local atmosphere which has uniform properties from hour to hour except that it undergoes a periodic fluctuation of surface pressure and whatever may be related to this, such as the heights and temperatures of pressure levels. For instance, let H be the height of the mean pressure level at which mesons are produced, and T be the temperature in the neighbourhood of that level. Then, according to Duperier (1949) a change of surface pressure ΔB is accompanied by a change of cosmic ray intensity ΔI such that

$$\Delta I = \mu\Delta B + \mu'\Delta H + \alpha\Delta T,$$

where μ, μ', and α are the mass, decay, and positive temperature coefficients respectively. The barometer coefficient β is then

$$\beta = \Delta I/\Delta B = \mu + \mu'\Delta H/\Delta B + \alpha\Delta T/\Delta B.$$

---

* Physics Department, University of Tasmania, Hobart.

In a local atmosphere in which $\Delta H/\Delta B$ and $\Delta T/\Delta B$ remain constant throughout the day, the changes $\beta\Delta B$ may be estimated by the use of the standard barometer coefficient β. The corrected cosmic ray intensities refer to a uniform atmosphere of constant surface pressure. Provided this reference pressure is chosen within the range for which the pressure dependence is linear, the difference between the pressure corrections $\beta\Delta B_x$ and $\beta\Delta B_y$ for two different hours will remain the same for any choice.

However, in a real atmosphere the coefficients $\Delta H/\Delta B$ and $\Delta T/\Delta B$ are not constant and the barometer coefficient will vary. Thus the pressure corrections at two different times will become $\beta_x\Delta B_x$ and $\beta_y\Delta B_y$. The difference between these obviously depends upon the choice of reference pressure, and this choice will therefore affect the form of the " corrected " daily intensity variation.

Variations in the relative intensities of the low and high energy primaries would also produce changes in the barometer coefficient, since this is a function of the energy of the particles arriving at sea-level (Fenton 1952).

In what follows it will be shown that there is evidence for a diurnal variation of the barometer coefficient and that one has to be very careful before ascribing diurnal variations remaining in the hourly cosmic ray intensities after correction, to variations in the primary radiation. This applies particularly to cases such as the present one, in which only sea-level meteorological data are available.

II. Results

A vertical counter telescope with a counting rate of approximately 80,000 counts/hr and an opening angle of 60°, measuring cosmic radiation which penetrates 10 cm of lead absorber, has been in operation at Hobart, Tas., since September 1953. Five 3-monthly groups of data have been analysed, from December 1953 to February 1955 inclusive. Within each group the days have been divided into six 4-hourly periods, and for each of these periods the mean intensity, surface pressure, and barometer coefficient have been obtained. Only complete days have been considered, and of these all days during which the passage of meteorological fronts occurred have been discarded.

The results, showing the daily variations of the barometer coefficient for each seasonal group, are displayed in Figure 1, in which the errors shown are the 95 per cent. fiducial limits. The last point on each graph refers to the same local time as the first point, but for the following day, so that it can be seen that not only are there periodic effects, but that quite large secular changes occur between the end points in some cases. As will be shown later, the large secular effect is due to the frequent passage of cold fronts, and the repeating pattern between cold fronts. Because of the errors of the end points, only a rough correction can be made for the secular effect but it appears that there are seasonal changes in the amplitude of the periodic variations and probably changes in phase as well.

Consider what happens when the observations are averaged over the year from December 1953 to November 1954 inclusive. Figures 2 (*a*), 2 (*b*), and 2 (*c*) show the variations of surface pressure, intensity, and barometer coefficient respectively. An obvious feature is that there is practically no secular change of intensity corresponding to the secular decrease in pressure. On the basis of the existence of a standard barometer coefficient one would expect an increase of intensity of 0·7 scaled counts at the end of the day for the fall in pressure of 0·83 mb. The secular decrease of the barometer coefficient is the cause of this apparent anomaly.

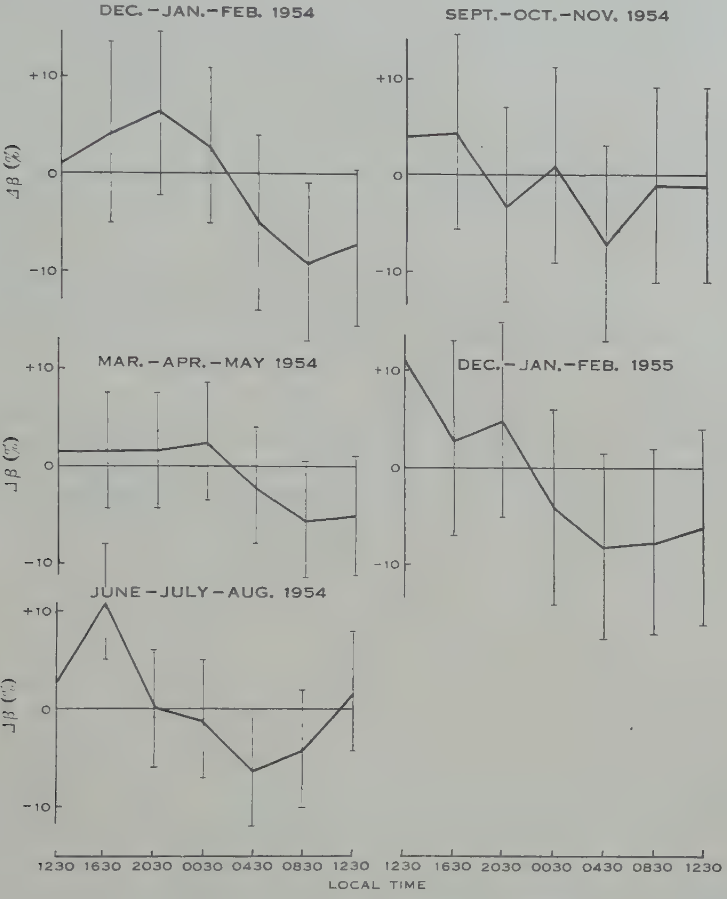


Fig. 1.—The diurnal variation of the barometer coefficient in 3-monthly groups. Errors shown are the 95 per cent. fiducial limits.

In Figure 2 (*d*) the full line shows the variation of intensity about the annual mean, corrected to an arbitrary pressure level using the standard coefficient, and adjusted for secular change. Alternatively correction could have been carried out using the observed coefficient appropriate to each 4-hr period and a reference pressure equal to the annual mean. There is no practical difference in the corrected values whichever method is used. Dashed lines show the effect of using the observed 4-hourly coefficients to correct the data to a pressure 15 mb above, 5 mb above, and 15 mb below the annual mean. The tails drawn are the 95 per cent. fiducial limits. There is a clear trend from a large amplitude at high to a small amplitude at low reference pressures. In fact, comparing the barometer coefficient variations, adjusted for secular change, with those of intensity corrected with the standard coefficient, it is interesting to notice that high barometer coefficients correspond to low intensities and vice versa. It is possible to choose a reference pressure at which the range of variation of the corrected intensity is within the statistical limits. The implication is that, for the year 1954, the cause of the barometer coefficient variations may have been the cause of the diurnal variations in the observed cosmic ray intensity.

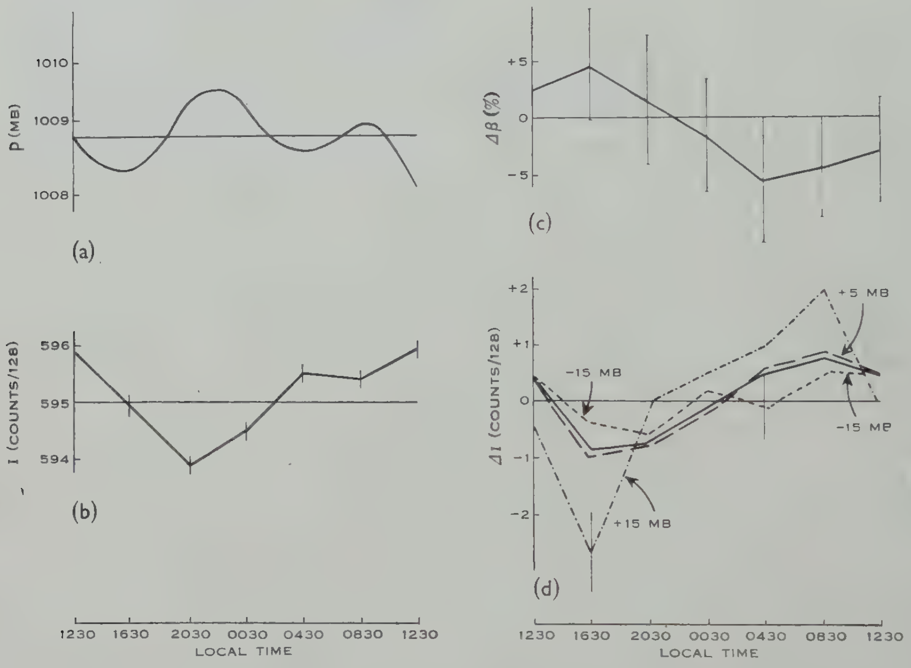


Fig. 2.—Diurnal variations averaged over the year December 1953 to November 1954 inclusive. Error tails show the 95 per cent. fiducial limits. (*a*) Surface pressure variation. (*b*) Variation of uncorrected intensity. (*c*) Variation of the barometer coefficient. (*d*) Variation of the corrected intensity. Full line—intensity variation corrected with a fixed long-term barometer coefficient to an arbitrary pressure level. The variation is practically identical if correction is made to the annual mean pressure using the observed 4-hourly coefficients. Dashed lines—the intensity variations corrected with the observed 4-hourly coefficients to a surface pressure 15 mb above, 5 mb above, and 15 mb below the annual mean pressure.

It should be pointed out that the apparent annual changes in amplitude and phase of the corrected intensity variations could be partly due to correction procedures.

III. Discussion

The periodicity of the coefficients may have a primary origin. Primary cosmic ray intensity variations which are not accompanied by changes in the energy spectrum cannot alone produce such a periodicity. If the energy spectrum does change, then it appears that it is such as to favour the arrival of high energy particles during the morning (from 3 a.m. to 3 p.m.) and lesser energies between 3 p.m. and 2 a.m. when the barometer coefficients are higher. Alternatively, the periodicities may be caused by changing atmospheric properties; for instance $\Delta H/\Delta B$ might be less between 3 a.m. and 3 p.m. than it is for the rest of the day. To distinguish between these possible influences a comparison should be made between the diurnal variations of intensity within a narrow energy band and the variation of the total hard component intensity. Then, if there is no appreciable diurnal variation of β within the selected narrow energy range by comparison with that for the total intensity it is likely that a change in primary spectrum is responsible for the overall β variations. Or, if the variations of β within the narrow energy band are the same as those for the whole spectrum, changing atmospheric properties must be producing the main effect.

Whatever may be the cause of the truly periodic variations, the secular decrease of the barometer coefficient is of atmospheric origin. It has come about in the following way. The data have been selected to exclude all days during which frontal passages occurred. In Tasmania, these are almost always cold, or occluded cold fronts. In other words, as far as possible all periods of change of air mass accompanied by rapid cooling have been eliminated. If they were not, there would be present in the data a number of brief periods containing abrupt increases of intensity of the order of 1 per cent., together with increases of the barometer coefficient of the order of 100 per cent., and it is quite impracticable to correct for these effects (Jacklyn 1954). Thus there remain the days during which the atmospheric temperature up to the production level for mesons is on the average increasing. Considering in this context only the mass effect and μ-meson decay, the barometer coefficient β is $\beta=\Delta I/\Delta B=\mu+\mu'\Delta H/\Delta B$. It has been shown (Jacklyn 1954) that $\Delta H/\Delta B$ is less in warm air masses than it is in cold air in the neighbourhood of fronts, so that gradual warming associated with a decrease of surface pressure should be accompanied by a decrease in the barometer coefficient. It is significant that all the seasonal groups of data show a secular decrease in pressure as well as a decrease in the barometer coefficient.

Corrections to the observed intensity variations for a purely secular change of β have periodic components due to the periodic variation of surface pressure. These do not vanish when the final linear adjustment is made for the secular change of corrected intensity.

IV. Conclusions

The consequences of the variations of the barometer coefficient for the study of cosmic ray diurnal variations may be summarized as follows:

(*a*) The data should be selected to exclude days during which frontal passages occur, otherwise large irregularities will obscure the relationship between cosmic ray intensity and surface pressure.

(*b*) Significant secular changes in the barometer coefficient should be expected. These are of atmospheric origin.

(*c*) If two sets of data, between which there is a large difference of the mean station level pressure, are to be compared under conditions of standard atmospheric pressure, then the diurnal variations of the barometer coefficient should be obtained and the data corrected using the observed coefficients. It should be noted that differences in mean station level pressure are involved, not differences in pressure due to altitude.

(*d*) Correction of data to an arbitrary standard atmospheric pressure does not entirely remove the diurnal variations produced by the atmosphere. The results from Hobart during 1954 suggest that at the time of maximum of the corrected diurnal variations of intensity there was a minimum barometer coefficient and at the time of minimum intensity there was a maximum barometer coefficient. Therefore, for this particular period the atmosphere may be largely responsible for the diurnal variations which have already been " corrected " for the atmospheric effect. These results, obtained during a period of low solar activity and presumably small primary intensity variations, serve to show that care must be exercised when attributing residual diurnal variations after correction to variations in the primary radiation.

(*e*) A changing primary spectrum and changing atmospheric properties may both contribute to the diurnal variations of intensity produced by the atmosphere. The study of intensity variations using a differential telescope should show which of these has the greater influence.

V. ACKNOWLEDGMENTS

The author wishes to acknowledge several valuable discussions of the draft manuscript with Professor A. L. McAulay, Dr. A. G. Fenton, and Mr. K. McCracken. Thanks are also due to Dr. Fenton for encouraging the project and correcting the proof, to Mrs. P. James for the numerous computations, and to Misses P. Hunt and J. Jackson for the typing. He is indebted to the Australian National Antarctic Research Expedition for the use of results from their counter telescope and to the Commonwealth Scientific and Industrial Research Organization for the funds which enabled the project to be undertaken.

VI. REFERENCES

DUPERIER, A. (1949).—*Proc. Phys. Soc. Lond.* **462** : 684–96.
FENTON, K. B. (1952).—Ph.D. Thesis, University of Tasmania.
JACKLYN, R. M. (1954).—*Aust. J. Phys.* **7** : 315–21.

WIND PROFILES OVER THE SEA AND THE DRAG AT THE SEA SURFACE

By E. L. Deacon,* P. A. Sheppard,† and E. K. Webb*

[*Manuscript received June 22, 1956*]

Summary

Wind speeds were measured at heights up to 13 m over the sea by means of anemometers mounted on a mast attached to the jib-boom of a small vessel and another on the foremast cross-trees. The vertical temperature gradient and air/sea temperature differences were also observed. On some occasions wind and temperature structures were explored up to 80–100 m by kite balloon.

A method of observation was devised whereby the effect of the ship's hull on the wind speeds was eliminated and from the corrected profiles surface drag coefficients were calculated. For neutral conditions and fetch of wind over the sea of 20–40 km it is found that the drag coefficient (for 10 m reference height) is about 0·0010 in light winds rising to about 0·0021 at wind speeds from 10 to 15 m/sec.

The drag coefficient is probably rather sensitive to stability variations, increasing particularly in the direction of instability, but the effects of variation in fetch of the wind over water are much less evident. In proximity to shallow water the drag coefficient increases appreciably.

I. Introduction

The importance to both meteorology and oceanography of a proper knowledge of the friction between atmosphere and ocean is evident, but observational difficulties have made progress slow and much uncertainty remains (see, for example, Francis 1954). Work over land surfaces has shown that the Prandtl logarithmic wind profile provides a secure basis for the estimation of the surface shearing stress under neutral conditions of stability but accurate measurement of wind profiles at representative exposures over the sea is no easy task. Fixed installations are normally limited to comparatively shallow water and the results are therefore not necessarily representative of the open ocean. Some studies of wind profiles over the sea are of doubtful significance for this reason and others have been vitiated by faulty exposure of the anemometers, as shown by Roll (1949). Erections in deeper water have to be so massive to resist storms that the disturbance to the wind field is considerable and to mount the anemometers beyond the perturbed region, when known, is still difficult. A ship also obstructs the flow to some considerable extent and its motion in a rough sea is a complicating factor but the advantage of being able to make observations in deep water away from the land is so great that it seemed worth while to try to find means of overcoming these difficulties. That this has been

* Division of Meteorological Physics, C.S.I.R.O., Aspendale, Vic.

† Imperial College of Science and Technology, London.

largely achieved it is hoped to show in the following account of observations of wind profiles made at sea in October 1955.

On some occasions wind and temperature observations were extended up to 80–100 m using a kite balloon and these results are presented here. Records were also taken of the turbulent fluctuations of wind and temperature using quick-response elements mounted at 13 m height but the analysis of these will be dealt with elsewhere.

II. Observational Procedure

The observations were made on the Fishery Research Vessel *Derwent Hunter*, a 70 ft Diesel-schooner of the Division of Fisheries and Oceanography, C.S.I.R.O., which for the duration of the trials (Oct. 6–26, 1955) was based on Portsea in Port Phillip Bay close inside the entrance (see Fig. 1). After an initial exploratory period in Port Phillip Bay observations were made in Bass Strait a few miles out from Port Phillip Heads whenever the wind had a southerly component. With northerly winds a suitable station in Port Phillip Bay permitted observations to be made with fetches of the wind over water of up to

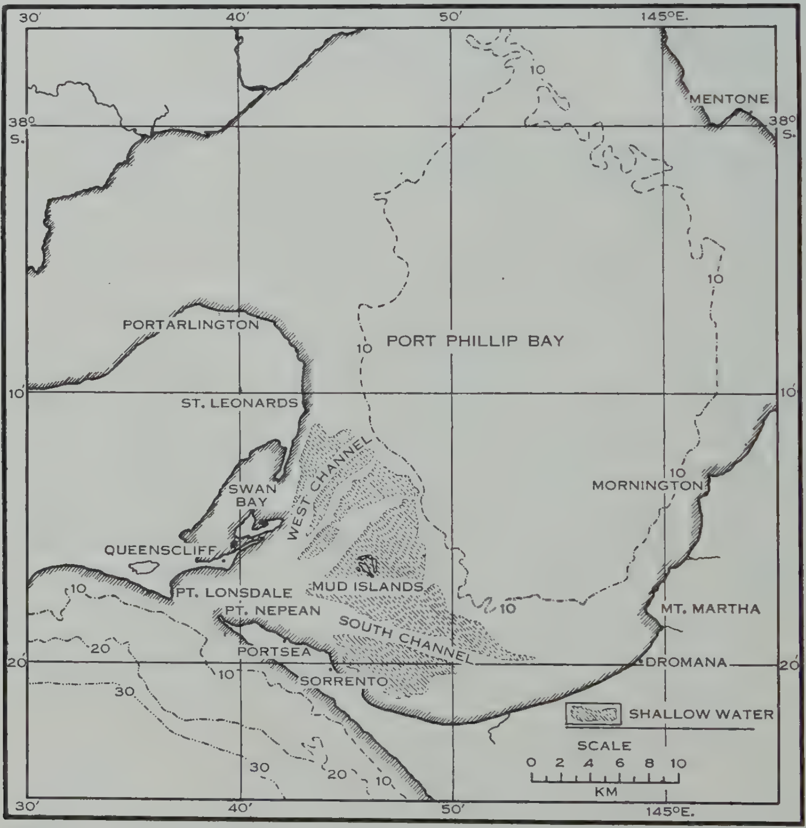
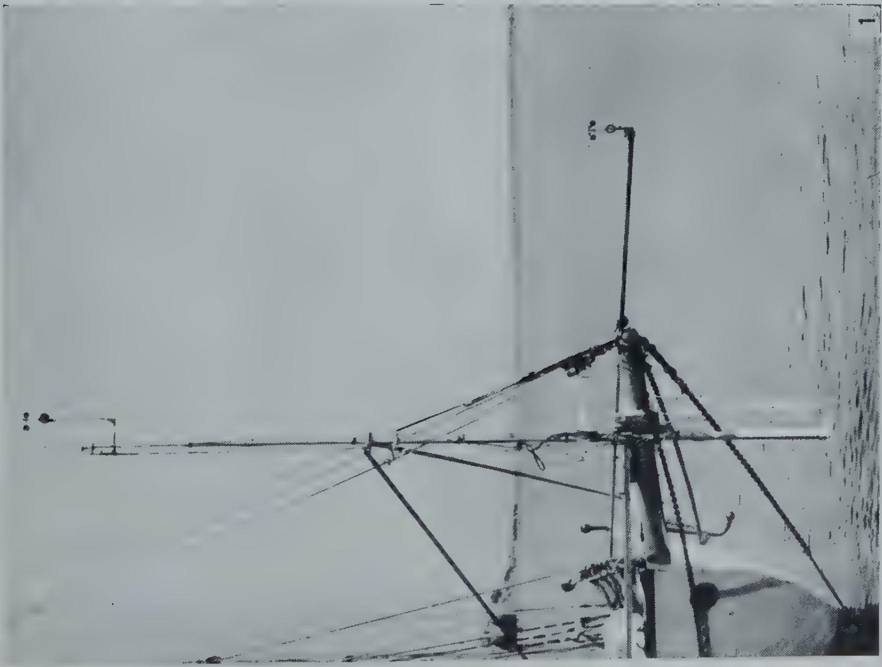
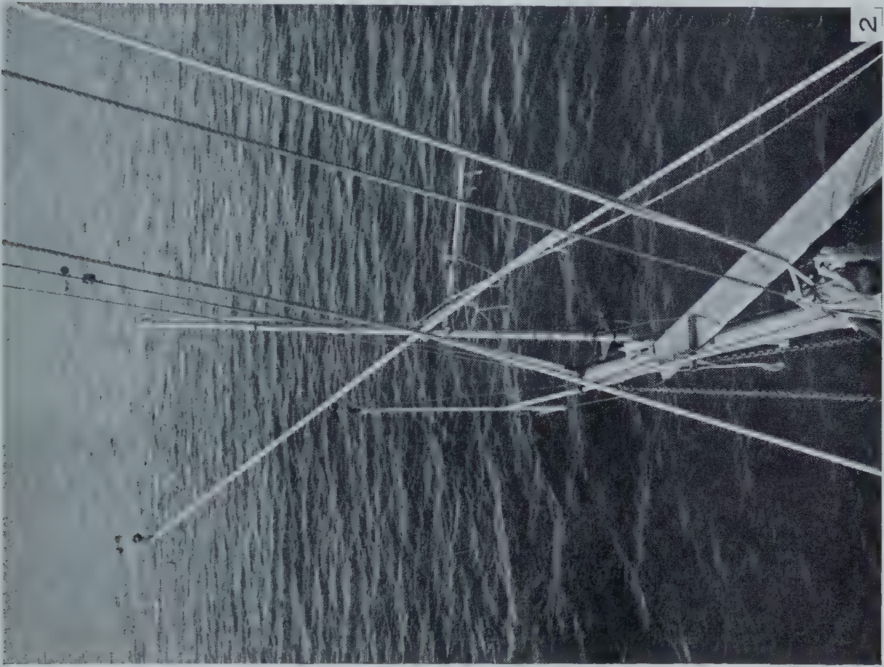


Fig. 1.—Area of observations. Depth contour lines are marked in fathoms.

WIND PROFILES OVER THE SEA AND THE DRAG AT THE SEA SURFACE



1



2

Fig. 1.—Bows of F.R.V. *Derwent Hunter* with anemometer mast fitted on the jib-boom.

Fig. 2.—The anemometer arrangement on October 13, 1955: looking down on the jib-boom from the foremast rigging.

50 km. Apart from an area of shallow water at the south-west end of the bay, which was avoided as far as possible, the general depth of water is 15–25 m.

An anemometer mast with arms was rigged on the jib-boom of the vessel at 2·3 m forward of the stem and carried several (up to five) anemometers at heights above mean water level from 1·5 to 6·4 m. The arrangement, which varied in detail during the course of the trials, is shown in Plate 1, Figure 1. A thermometer element was also carried on this mast generally at a height of 4 m except for the first 2 days, when the height was 3 m. Another anemometer was mounted on the foremast cross-trees at a height of 12·95 m together with a thermometer element at 12·6 m. This upper anemometer level is hereafter referred to as 13 m although 12·95 m has been used in the calculations. At various times anemometers were carried at positions mainly about 2·7 m forward of the anemometer mast by means of booms of 1½ in. steel pipe. These were used to get some information on the extent to which wind speeds at the anemometer mast were influenced by the presence of the ship's hull. Plate 1, Figure 2, a photograph from the foremast rigging looking down on the bow, shows the arrangement on October 13. The highest anemometer (6·4 m) in this figure is on a sub-mast mounted alongside the anemometer mast proper using a spring-loaded catch device, a design which greatly facilitated removing the instrument. An anemometer at 5·8 m is to be seen on a slanting boom projecting forward and to the port side. Two more anemometers (1·95 and 3·95 m) are carried by the L-shaped piece on the end of a boom extending the jib-boom. The 4 m arm on the anemometer mast is projecting to starboard; when an anemometer was mounted on this it was swung forward to be at 30° to the fore-and-aft line.

During observation periods the ship steamed slowly into wind at the lowest speed for sufficient steerage way to be maintained. The relative wind direction was generally within 10° of the fore-and-aft line and was but rarely as much as 15° on the bow. The anemometer readings together with measurement of the speed of the ship relative to the water enabled wind speeds relative to the sea surface to be found. The ship's speed, usually around 1–2 knots, was found by timing the passage of floating objects (orange peel) between two points along the ship's side. Orange peel floats nearly totally submerged and the colour makes it easy to observe. The peel was flung about 20 ft out from the ship's side to avoid effects of the propeller intake, etc. An average of 7 determinations were made per 30 min run. Nearly all runs were of 30 min duration after the first 4 days during which the observation period was mostly 10 min.

To investigate the effect of the ship's hull on the wind profile some runs (" speed runs ") were made at full speed (7½ knots) into wind for comparison with normal runs. These were made on occasions of moderate to low wind speed.

Sea and air temperatures, the latter at about 1·5 m above deck level, were observed during each run and the temperature difference between 4 and 12·6 m was recorded on most days. On some days the wind and temperature structure for greater heights (up to about 100 m) was observed using an instrument lifted by a " Kytoon ".

III. Instrumental

(*a*) *The Anemometers*

The Sheppard type sensitive cup anemometers employed an electric contacting arrangement similar to that described by Crawford (1951) but the contacting rate was only half as great (100 counts/min$\sim$10 m/sec). Telephone message registers (6 V) were used for counting. The anemometers were calibrated in a wind-tunnel immediately before and after the trials and were also checked on October 19, about half-way through. Only one anemometer showed any significant change, about 1·5 per cent., and for this the initial calibration was used up to October 19 and the final thereafter.

The anemometers were taken down at the end of each day's work and checked over. When replaced at the beginning of the next day, interchanges were made so as, in the mean, to minimize errors from calibration inaccuracies.

(*b*) *Temperature Recording System*

Temperatures were recorded using thermistor elements (S.T.C. type F2311/300) in a Wheatstone bridge circuit, the output of which was applied to an Evershed & Vignoles recording milliammeter (2 mA full scale, 1250 Ω) via a Sunvic D.C. amplifier. The thermistors were switched in turn into the circuit to give samples, for a minute or two each, of temperatures in the sea and at 4 m, 12·6 m, and " Kytoon " height. The bridge was approximately balanced each time by the resistance box in one arm of the bridge. The average of the recorded trace, estimated by eye, provided simply a small correction to the resistance box reading. Once each day the thermistors were calibrated at a bath temperature equal to about the mean temperature for the day's runs. The 4 and 12·6 m elements were calibrated together in the same bath to enable the temperature difference between these heights to be obtained with the minimum error.

To improve the accuracy of determining the 12·6 : 4 m temperature differences, the circuit was changed to the differential type on October 17 and, by omitting the sea element, the sensitivity could be increased to 2½ °F full scale. The slight change of zero with temperature due to imperfect matching of thermistors was determined each day with pairs of elements in the same bath at several temperatures spaced about 1 °F.

For air temperature measurement each thermistor was mounted in a radiation shield consisting of a pair of concentric chromium-plated cylinders. The shield was open at both ends and pointed forwards so that ventilation was automatically provided when the ship was headed into wind. The sea temperature element was placed in a sealed metal tube mounted on a weighted metal frame hung in the sea at a depth of about 0·5 m. In addition to the thermistor measurements, sea and air temperatures were taken on all occasions using mercury-in-glass thermometers. Except on rough days Assmann psychrometer readings were taken with the instrument suspended from the forestay. On rough days a whirling psychrometer was used amidships with the instrument held out to give as good an exposure as possible. Sea temperatures were taken with a bucket consisting of a large, wide-mouthed Dewar flask mounted in sponge rubber in a protective cylinder.

(c) *The Captive Balloon-sonde*

A combined anemometer and temperature element was made for the balloon work. The construction, mainly in bright anodized aluminium, is sufficiently clear from Figure 2. The vane-wheel, 6·7 cm in diameter, operated an electric contact through a worm and wheel reduction gearing. The thermistor was housed in a lighter version of the double-tubular shield used on the ship. The instrument was inserted in the balloon cord some 15 m below the balloon—a Darex " Kytoon "—and electrical connection was by means of four P.V.C. covered copper wires, 0·038 cm diameter, twisted together and tied at 2 m intervals to the nylon cord. Weights were 7 g/m for the combined cable and 350 g for the instrument.

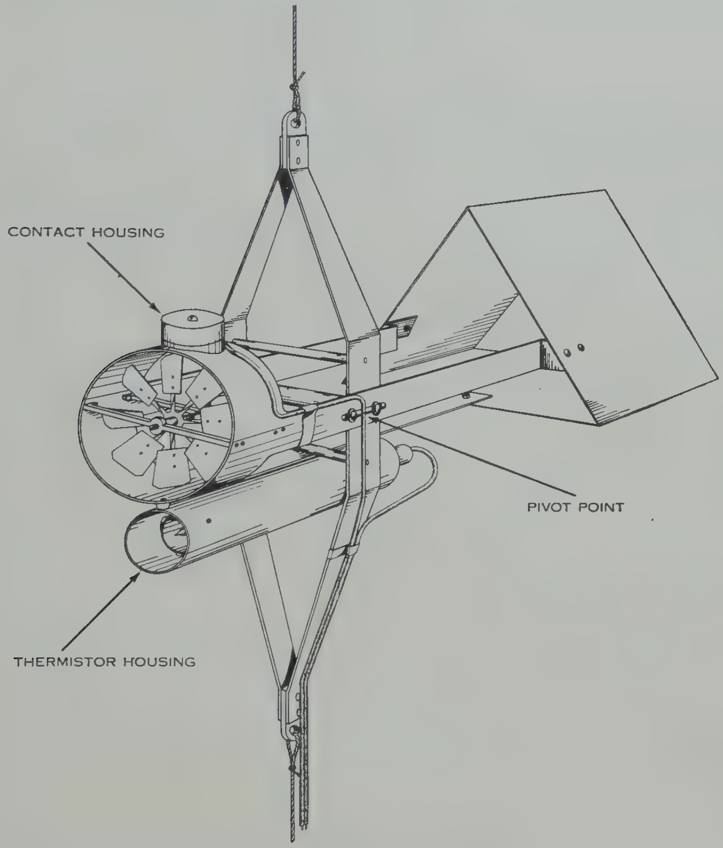


Fig. 2.—The captive balloon-sonde.

At first " Kytoons " of 2·3 m³ capacity were used and, with the better of the two available,* heights of over 100 m were attainable. After this was lost in a strong wind, a smaller " Kytoon " (1·1 m³ capacity) was used at heights up to some 80 m. Readings were taken at each of several heights in turn and, assuming a catenary form for the cable, these heights were subsequently calculated from the cable length, the elevation angle of the instrument, and angle of the cable to the horizontal where it left the winch drum, the last two being averaged over several clinometer readings for each recording.

---

* Few balloon results were obtained in the first half of the trials owing to the first large " Kytoon " flying badly, probably owing to some asymmetry.

The cable proved to be a weak feature of this equipment as the wires were rather delicate and the smallest crack in the insulation is serious in the presence of a salt film. The insulation was easily damaged at the points of attachment to the nylon cord.

IV. EFFECTS OF THE SHIP ON THE WIND PROFILES

(a) *Obstruction of the Air Flow by the Ship's Hull*

The ship influences the wind field in its vicinity and, despite the fact that the anemometers were rigged well forward on a small ship with moderately fine lines and no raised forecastle, it was considered likely that there would still be an effect too large to be neglected.* This was found to be the case during the trials, when anemometers on booms were rigged forward of those on the anemometer mast. For example, 31 runs were made with an anemometer at 5·85 m carried 2·5 m forward of the 6·4 m anemometer. Using the final wind profiles to correct for the small height difference, it is found that the wind velocity at the forward position was $1\cdot 2\pm 0\cdot 3$ per cent. greater than at the same level on the anemometer mast. A similar difference (0·9 per cent., mean of 7 runs) is found at the 4 m level.

The comparison of results of normal and speed runs is used to derive corrections for hull effect in the following manner. It is assumed that the relative wind speed at an anemometer position at height z is increased due to the presence of the hull by a factor f_z which is taken to be independent of relative wind speed. For simplicity the 13 m value of f is taken as unity. This gives corrections relative to 13 m, which is all that is required.

For a standard run at ship speed S_1 we find the recorded wind ratio, R_1 (z : 13 m) to be related to the true wind speed relative to the surface, $u_1(z)$, as follows:

$$R_1=\frac{f_z\{u_1(z)+S_1\}-S_1}{u_1(13)}.$$

For a speed run at ship speed $S_2>S_1$

$$R_2=\frac{f_z\{u_2(z)+S_2\}-S_2}{u_2(13)},$$

and the difference of these ratios gives

$$R_2-R_1=f_z\left\{\frac{u_2(z)}{u_2(13)}-\frac{u_1(z)}{u_1(13)}\right\}+(f_z-1)\left\{\frac{S_2}{u_2(13)}-\frac{S_1}{u_1(13)}\right\}.$$

---

* Sails were not set at any time in the trials.

The first term on the right-hand side will be zero if no changes in drag coefficient of the sea surface or in stability have taken place between the runs. Assuming this to be so, then

$$f_z = 1 + \frac{R_2 - R_1}{S_2/u_2(13) - S_1/u_1(13)}. \quad \ldots\ldots\ldots\ldots\ldots (1)$$

If there is but a small change in wind speed between the runs compared, then equation (1) is approximated by

$$f_z = 1 + \frac{(R_2 - R_1)u(13)}{S_2 - S_1},$$

which shows that for accurate evaluation of f_z by this method the change in ship speed between the runs must be of reasonable magnitude compared with the 13 m wind speed. Very light winds, however, are not generally favourable as they are too often variable in strength and direction.

The values of f calculated from equation (1) give moderately consistent results, as may be seen from the following percentage effects, $100(f_z - 1)$, for the 6·4 m position as a result of 10 speed runs on different occasions :

−0·3, −3·0, −3·6, −1·6, −2·8, −3·2, −4·3, −3·0, −3·9, −0·8

Mean −2·65 per cent. with standard error 0·42 per cent.

The corresponding mean for the 5·85 m boom anemometer from four speed runs is −1·2 per cent. so these results are in agreement, within the limits of observational error, with the fact that this boom anemometer registered a wind speed 1·2±0·3 per cent. higher than found for the same level at the anemometer mast position.

In this way the corrections for hull effect given in Table 1 were compiled.

TABLE 1

HULL EFFECT CORRECTIONS TO OBSERVED RELATIVE WIND SPEEDS

Anemometer Position	No. of Speed Run Comparisons	Positive Correction (%)
6·4 m mast	10	2·7±0·4
5·85 m boom	4	1·4
4 m mast (on 0·6 m arm) ..	2	4·0
4 m mast (on 1·5 m arm) ..	9	3·4±0·6
3·95 and 3·68 m boom ..	4	2·5
3 m mast (0·6 m arm) ..	1	7·5
2·67 m and 1·95 m boom ..	6	3·9±0·7
1·5 m mast (0·6 m arm) ..	3	12·5
1·5 m mast (1·5 m arm) ..	4	9·5

It is interesting to note that even 5 m forward of the stem of the vessel the effect of the hull is felt to the extent of 2 or 3 per cent. Neglect of this effect would lead to vertical wind gradients seriously in error. Had the rather large magnitude of the hull effect been fully appreciated early in the trials more speed runs would have been made with a consequent gain in precision of the corrections.

The wind speeds shown in the table of results (Appendix I, Table 5) are those found after adding the above corrections to the observed relative wind speeds and then subtracting the ship speed S.

(*b*) *The Effect of Rolling*

The lateral wind component caused by rolling results in the recorded wind speed (U_r) being greater than the relative wind speed which it is desired to measure. Here wind speeds relative to the ship are being considered. When the ship is running head to wind we have, for a harmonic rolling motion,

$$U_r^2 = U^2 + (2\pi h\alpha/T)^2 \sin^2 \omega t,$$

where U=actual relative wind velocity (assumed constant),
h=height of anemometer above axis of roll,
α=rolling amplitude, radians,
T=rolling period,
$\omega=2\pi/T$,
t=time.

Taking mean values over an extended period (i.e. $t \gg T$)

$$\bar{U}_r/U = (2/\pi)\int_0^{\frac{1}{2}\pi} \sqrt{(1+b^2 \sin^2 \omega t)}\,\mathrm{d}(\omega t),$$

where $b=2\pi h\alpha/TU$. For the present purpose we are only concerned with solutions for $b<0\cdot 3$ so, expanding in series before integration and neglecting terms containing the fourth and higher powers of b, we find

$$\bar{U}_r/U = 1+\tfrac{1}{4}b^2. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (2)$$

Precise observations of mean rolling amplitude were not made on these trials but it was rarely as much as 7° and the higher values were confined to some runs in Bass Strait when the direction of the swell was different from that of the wind. Values of drag coefficient have not been evaluated for such occasions whenever the wind speed was less than 8 m/sec (i.e. October 10 and 26) as the rolling effect would then be rather large.

For a relative wind speed of 10 m/sec and a rolling amplitude of 7° with the observed period of 4 sec, we find that the observed 13 : 4 m winds ratio is larger than the true value by 1·2 per cent. and this may be shown to correspond to an error of about 17 per cent. in the calculated shearing stress. The drag coefficients for Bass Strait presented later are therefore up to some 10 or 15 per cent. too large but those for Port Phillip Bay should be in error by less than 5 per cent. as rolling there was in general quite slight, when head to wind, owing to the absence of swell.

Fore-and-aft motions of the anemometers caused by pitching and surging do not give rise to error in the wind profiles but there will be an effect of the varying attitude of the anemometers caused by pitching. This has been neglected as the angle of pitch is small and the attitude effect should not vary much with height.

(c) *Effect of Variation in Height of Anemometers above the Water Surface*

In the course of an observation the height of the anemometers above the water fluctuates considerably about the mean owing to the waves and the pitching of the vessel and, as the variation of wind speed with height is not linear, an error is introduced in attributing the mean wind speed indicated by a given anemometer to the mean height of that instrument. The form of the wind profiles is such that the error will be larger at the lower levels. Even were data available for the amplitude of the fluctuation in height, a precise calculation of this error would require knowledge of the variations in wind speed over the waves as between the crests and troughs. It is useful, however, to make an estimate of the error on the basis of the mean wind profile. The neutral profile is

$$u=(u_*/k)\ln(z/z_0),$$

where u_*=the friction velocity,

z_0=roughness parameter,

k=the Kármán constant,

and the anemometer is assumed to be subject to a harmonic variation in height $A\sin\Phi t$. The mean wind speed recorded by the anemometer ($\bar{u}_r$) is therefore

$$\bar{u}_r=\overline{(u_*/k)\ln[(z+A\sin\Phi t)/z_0]}.$$

Provided that the period of the vertical oscillation is small compared with the observation period and $A<z$, this can be approximated by

$$\bar{u}_r=(u_*/k)[\ln(z/z_0)-\tfrac{1}{4}A^2/z^2]. \quad \ldots\ldots\ldots\ldots\ldots (3)$$

From this it follows that

$$\bar{u}_r(z_2)-\bar{u}_r(z_1)=\left(\frac{u_*}{k}\right)\ln(z_2/z_1)\left[1-\frac{A^2(z_2^{-2}-z_1^{-2})}{4\ln(z_2/z_1)}\right].$$

In the present work u_* was evaluated from the wind speed difference between 4 and 13 m, and it follows that the error in u_* is then $1\cdot20A^2$ per cent. (A measured in metres), the sense of the error being such that the observed values are too great. It was noted during the runs made in the rougher conditions that the foot of the anemometer mast (0·9 m above mean water level) was fairly frequently submerged in the wave crests but the average value of A in these runs was generally rather less than 0·9 m. On a very few runs A may have been rather more than 1 m but 1·3 m is an upper limit so the maximum error in u_* on this score is about 2 per cent. corresponding to a 4 per cent. error in drag coefficient. It is likely that the above treatment based on the *mean* wind profile gives an overestimate of this effect, as any tendency for the wind over the wave crests to be greater than that in the troughs would entail smaller errors.

V. Calculation of the Shearing Stress from the Wind Profile

The wind speeds given in the table of results (Appendix I, Table 5) were plotted against the logarithm of height and from these graphs the values of $u_{13}-u_4$ were approximately evaluated and used with the corresponding potential temperature difference† $\theta_{13}-\theta_4$ to give layer Richardson numbers $Ri\ (13:4)$ for each run from

$$Ri\ (13:4)=\frac{g}{\theta_4}\,\frac{\Delta\theta\Delta z}{(\Delta u)^2},$$

where the differences are for the height interval 4–13 m. In the following Ri should be understood to be $Ri\ (13:4)$ unless stated otherwise. On two days potential temperature differences were not observed and values have been estimated from the air/sea temperature differences in the light of the rest of the data (see Section VII). It was immediately apparent that for many of the runs the value of Ri was significantly different from zero and that stability effects on the form of the profiles would have to be taken into account.

The observations could not all be treated in uniform fashion over the whole stability range ; it was found necessary to adopt two systems, one for the range Ri $-0\cdot025$ to $+0\cdot10$ and the other for the observations at greater instability. A few profiles for $Ri>0\cdot10$ have not been used to evaluate shearing stresses as no reliable method is available.

For Ri between $-0\cdot025$ and $+0\cdot10$ it is considered in the light of other observational work (Sverdrup 1936 ; Pasquill 1949 ; Deacon 1953, 1955 ; Rider 1954) that the Rossby-Montgomery (1935) formulation of the effect of stability provides the most acceptable basis for the estimation of surface shearing stress (τ_0) but a considerable element of uncertainty remains. The relationship is

$$\frac{u_*}{z\partial u/\partial z}=\frac{k}{\sqrt{\{1+\sigma Ri(z)\}}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

where $u_*=(\tau_0/\rho)^{\frac{1}{2}}$,
ρ=air density,
k=the Kármán constant (taken to be $0\cdot40$),
σ=the Rossby stability constant.

The reason why some wind-profile observations (e.g. Deacon 1953) under moderately strong stability ($0\cdot05<Ri<0\cdot10$) have indicated a failure of equation (4) is now considered to be a consequence of the shearing stress having been estimated on the assumption that it is proportional to the square of the wind velocity close above the surface roughness elements. This is only true if the Reynolds number of the flow is sufficiently large and there is evidence now that this is often not the case (on land) under these conditions. This matter is dealt with in a note to be published by one of us (E.L.D.) elsewhere.

The value of σ to be used in (4) is not yet known with accuracy. Sverdrup (1936) found $\sigma\sim11$ and Deacon (1953) about 6 or 7, but in both these studies τ_0 was not measured but estimated using the roughness parameter z_0 indicated by the neutral wind profiles. The values of σ found in this way are very sensitive to errors in z_0 ; the above difference between 7 and 11 corresponds to the small change in z_0 from $0\cdot25$ to $0\cdot30$ cm. Rider's (1954) direct measurements of τ_0

---

† The upper temperature element was at a height of 12·6 m but this has been taken to be close enough to 13 m for the difference to be neglected.

together with wind and temperature profiles provide, at present, the best material for an evaluation of σ. Using the values of $\partial u/\partial z$, Ri, and τ_0 given by Rider (op. cit., Tables 2 and 3) the value of $\sigma=8 \cdot 8$ was found from the least squares fit of equation (4) to the data in the range $-0 \cdot 025 < Ri < 0 \cdot 10$. In the reduction of the present data the value $\sigma=9$ is employed but it is apparent that this value is still uncertain by 1 or 2 units.

The application of equation (4) to our data in order to obtain u_* presents some difficulty. The equation has no established integral form, while the scatter of points in the wind profiles and the absence of detailed temperature profiles make a direct application of equation (4) impossible. We have proceeded as follows:

(i) The wind and temperature profiles between 4 and 13 m were assumed similar in form and to be representable by the interpolation formulae

$$\frac{\partial u}{\partial z}, \frac{\partial \theta}{\partial z} \propto z^{-\beta}. \qquad (5)$$

(ii) The assumptions (i) enable equation (4) to be integrated and lead to

$$\beta=\frac{1+\sigma Ri(z)}{1+1 \cdot 5\sigma Ri(z)}, \qquad (6)$$

and

$$u_*^2=F_1(u_{13}-u_4)^2, \qquad (7)$$

where F_1 is a function of $Ri(13:4)$ which has been calculated and is given in Table 2. The detailed procedure leading to Table 2 is given in Appendix II.

TABLE 2

VALUES OF F_1 FOR USE IN CALCULATING u_*^2 FROM THE 13 M—4 M WIND SPEED DIFFERENCE USING EQUATION (7)

Ri (13 : 4)	—0·02	—0·01	0	+0·02	+0·04	+0·06	+0·08	+0·10
$F_1 \times 10^3$	1·42	1·27	1·16	0·98	0·86	0·78	0·68	0·62

The problem is now to find the best value of $u_{13}-u_4$ from the wind data and so also of F_1 in order to obtain u_* from equation (7). The following steps were taken to this end.

(iii) Initial estimates of $u_{13}-u_4$ and hence of $Ri(13:4)$ were made (as already mentioned) and β calculated from equation (6). A theoretical profile was then prepared passing through the observed u_{13} and the value of u_4 corresponding to the wind difference.

(iv) This theoretical curve was slightly adjusted on the $u:\log z$ diagram until, maintaining the curvature unchanged, it gave a good fit to the observations and still ran through the 13 m point.

(v) A revised value of $u_{13}-u_4$ was then available from which was obtained a revised $Ri(13:4)$. The appropriate value of F_1 together with the wind difference then gave u_*^2 from equation (7).

(vi) From u_*^2 and u_{10} read from the fitted wind profile the drag coefficient $c_{10}=u_*^2/u_{10}^2$ was obtained. The values of c_{10} so derived are given in Table 5 of Appendix I.

It will be noted from step (iv) that, for each profile, the 13 m observation was accepted as correct and the lower level observations were utilized in obtaining an estimate of u_4. They were therefore given less individual weight than the 13 m value. This system was adopted, partly in order that the profile fitting should be as objective a process as possible, but also because it was appropriate to give rather less weight to the low-level observations in view of the much bigger hull effect at low levels. As all six anemometers took turns at 13 m systematic calibration errors at this height were excluded.

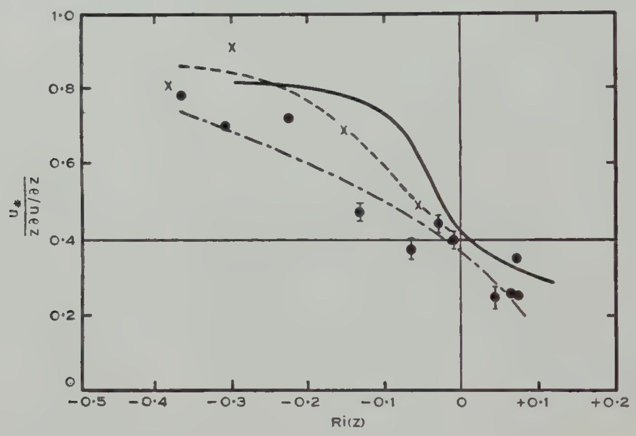


Fig. 3.—Variation of $\dfrac{u_*}{(z\partial u/\partial z)}$ with Richardson number.

——— Rider's (1954) mean curve.

●-·-· Observations by eddy correlation method (Deacon 1955).

×---- From wind profile data given by Deacon (1953) with $k=0\cdot40$.

The Rossby equation fails under unstable conditions when the instability exceeds a relatively small amount owing to the onset at about $Ri=-0\cdot03$ of a regime of effectively free convection (Priestley 1955). The observational evidence on the behaviour of $\overset{*}{K}\equiv u_*/(z\partial u/\partial z)$ at negative Richardson numbers is summarized in Figure 3, from which it is apparent that knowledge is still very imperfect in this range. There is some measure of agreement that $\overset{*}{K}$ tends to a value of about $0\cdot8$ at very large instability but in the neighbourhood of $Ri=-0\cdot1$ there is much uncertainty. It is considered that Rider's data are probably the most reliable of those at present available and accordingly they are employed in the reduction of our data for $Ri<-0\cdot025$, but the uncertainty in this reduction should be borne in mind. Rider's observations of u_* and of wind speed and temperature at heights of 50 and 150 cm are used to give the factor F_2 (see Table 3) as a function of $Ri(3a:a)$ in

$$u_*^2 = F_2(u_{3a} - u_a)^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

and on grounds of profile similarity it is assumed that this may be applied to our data with $a=4$ m.

The factors of Table 3 are applied to the wind speed differences $u_{12}-u_4$ read from the faired profiles.

TABLE 3

FACTORS FOR THE EVALUATION OF u_*^2 FROM WIND PROFILES AT MODERATE TO STRONG INSTABILITY USING EQUATION (8)

$Ri\ (3a:a)$	—0·025	—0·04	—0·06	—0·08	—0·10	—0·15	—0·20	—0·25
$F_2 \times 10^3$	1·97	2·7	3·3	3·7	3·9	4·5	5·0	5·4

VI. The Wind Profile Results and Discussion

(a) *The Wind Profiles and Drag Coefficients*

The wind velocities after correction for hull effects and ship speed are given in the table of results (Appendix I, Table 5) together with other relevant data and values of the drag coefficient $c_{10} \equiv u_*^2/u_{10}^2$ calculated as indicated in Section V.

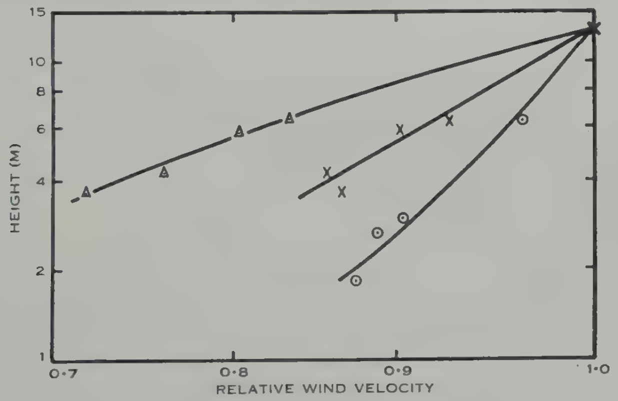


Fig. 4.—Specimen wind profiles. These examples are rather better than average but a greater part of the material is of similar quality.
⊙ run No. 17*a*, $Ri(13:4) = -0 \cdot 07$; × run No. 106, $Ri(13:4) = 0 \cdot 01$; △ run No. 103, $Ri(13:4) = 0 \cdot 18$.

Some specimen profiles at various degrees of atmospheric stability are shown in Figure 4. Approximately 75 per cent. of the profiles can be classified as good or fairly good while the remainder show more scatter of the points but are still usable. There is no evidence of any systematic departure of the profiles from the expected forms, i.e. linear on the $u:\log z$ plotting under neutral conditions and somewhat curved, but in opposite senses, under stable and unstable conditions.

The drag coefficients c_{10} are considered in relation to the 10 m wind speed in three groups: (i) the Bass Strait results, (ii) the Port Phillip Bay results, (iii) the observations for neutral and near-neutral conditions for both locations.

(i) *Bass Strait Results.*—These results are shown in Figure 5. The fetch of the wind over sea was from 20 to 40 km (average 27 km) for the observations at wind speeds above 8 m/sec and very large for the remainder. No great weight can be attached to the squares for which $Ri(13:4) < -0\cdot04$, owing to uncertainties in the calculation of u_* under such conditions.

The full line in Figure 5 has been drawn to represent, rather tentatively, the variation of c_{10} with wind speed under neutral conditions. There are no near-neutral results between 6 and 10 m/sec so here the curve has been drawn below the points for unstable conditions and mainly above those for stable stratification.

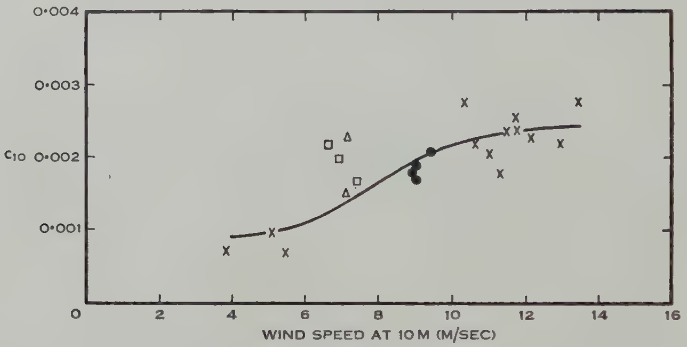


Fig. 5.—Drag coefficient related to wind speed; Bass Strait observations. The curve is drawn for neutral conditions. Range of $Ri(13:4)$: □ $<-0\cdot04$; △ $-0\cdot02$ to $-0\cdot04$; × $-0\cdot019$ to $+0\cdot029$; ● $+0\cdot03$ to $+0\cdot06$.

The three points for near-neutral conditions and $u_{10} < 6$ m/sec (runs 15–17) were obtained on an occasion when the sea and air temperatures differed by only a fraction of 1 °F and $\theta_{13} - \theta_3$ was zero within the limits of observational error. The three profiles are shown in Figure 6 with straight lines fitted to the values by least squares. There appears to be little doubt that on this occasion the drag coefficient was distinctly low as compared with other occasions of neutral stability but strong wind. The value of c_{10} for an aerodynamically smooth surface is expected from laboratory investigations to be about $0\cdot0009$ at 5 m/sec, a value close to that given, on a larger scale, by Van Dorn's (1953) measurements on a pond which had detergent applied to the surface to reduce wave formation.

The 10 near-neutral observations at wind speeds between 10 and 14 m/sec give a mean $c_{10} = 0\cdot00235$ with a standard error of $0\cdot00010$. But the motion of the ship under these conditions was such that this mean value is probably about 10 per cent. too large, mostly because of rolling but partly owing to oscillation in height of the anemometers, so the true value of c_{10} should be fairly close to $0\cdot0021$; the corresponding value of the roughness parameter is $0\cdot15$ cm.

(ii) *Port Phillip Bay Results.*—Altogether 58 values of drag coefficient are available for Port Phillip Bay. For these the fetch over water varied from 11 to 50 km, but in a considerable number of cases the first part of the fetch was over the shallow water above the sand banks in the south-western part of the Bay (see Fig. 1) where the character of the sea surface was observed to be very different from that over the deeper parts of the Bay (15–25 m). In view of this, fetch over the deeper water is also given in Table 5 of Appendix 1. Fortunately, the transition from shallow to deep water is sharp, so there was no difficulty in assigning values to the fetch over deep water. The correlation coefficient between drag coefficient and the logarithm of fetch over deep water is $-0 \cdot 37$ (58 values) while the partial coefficient after eliminating $Ri(13:4)$ is $-0 \cdot 34$, a value significant at the 1 per cent. level. The 39 observations for $u_{10} > 7$ m/sec give a similar result. In Figure 7 mean values after grouping for fetch are shown plotted against fetch over deep water. It seems from this that the drag coefficient may be relatively independent of fetch if this exceeds about 3 km over deep water but at smaller fetches (all to leeward of shallow water) where the surface is marked by short rather steep waves in a confused pattern, the drag coefficient rises to a rather higher value.

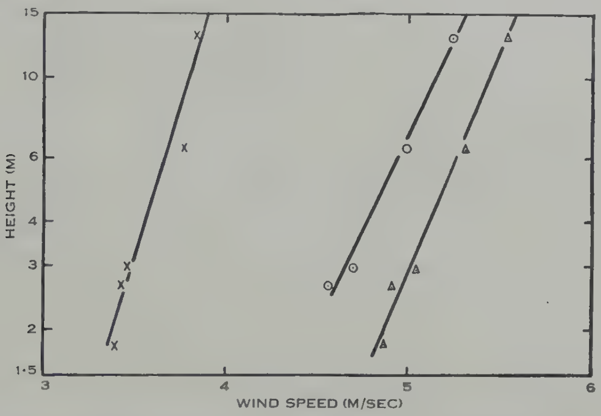


Fig. 6.—Wind profiles for conditions of neutral stability and light wind ; Bass Strait.
⊙ Run 14. × Run 15. △ Run 16.

The results for fetches over deep water of 5 km or more are shown in Figure 8. The observations are too few, considering the extensive range of stability and the uncertainty in interpretation of profiles under extreme conditions, for much more to be said than that the general level of the Bay values is much the same as for Bass Strait.

(iii) *Neutral Conditions* ; *Both Locations.*—As the near-neutral results are the most reliable they provide the best material for a comparison of the drag coefficients for Port Phillip Bay and Bass Strait. They are shown for both locations plotted together in Figure 9 in which the line drawn is from Figure 5. For the Bay observations the least fetch over deep water is 6 km. The two sets of results appear to accord reasonably well bearing in mind that the Strait values for high winds are about 10 per cent. too high owing to ship motion, whereas the Bay values are much less affected by these factors.

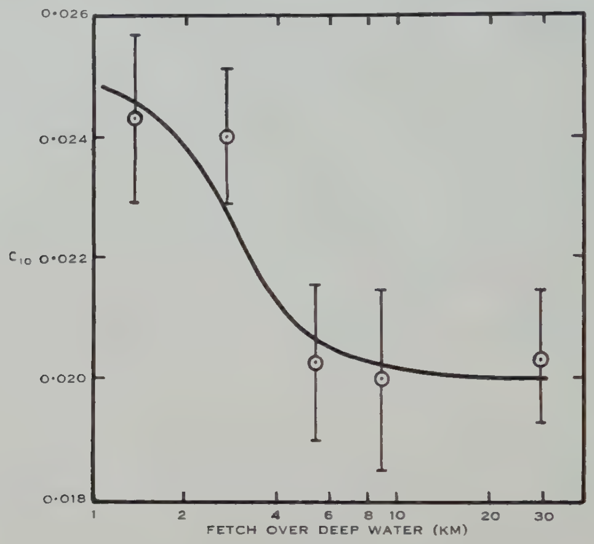


Fig. 7.—The rise in drag coefficient in proximity to shoal water. Observations in Port Phillip Bay.

Two near-neutral observations for Port Phillip Bay (runs 86 and 87—not shown in Fig. 9) are worthy of separate mention as they were made in a confused sea at a time when the wind was still changing direction appreciably after a thunder squall. The drag coefficients of 0·0021 and 0·0017 for 10 m

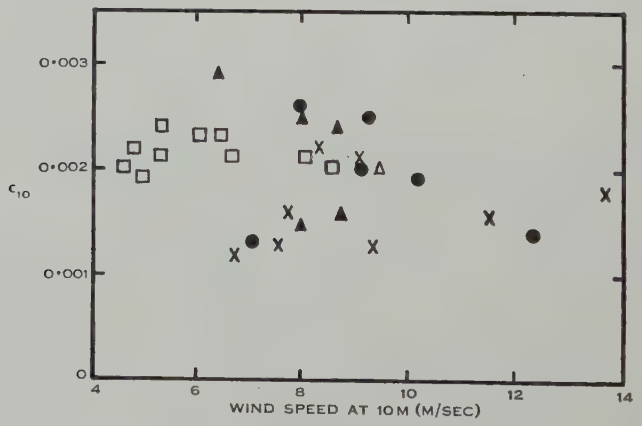


Fig. 8.—Drag coefficients from observations in Port Phillip Bay when the fetch of wind over deep water was 5 km or more. □ *Ri* <—0·04 ; △ *Ri* —0·02 to —0·04 ; × *Ri* —0·019 to +0·029 ; ● *Ri* +0·030 to +0·059 ; ▲ *Ri* +0·06 to +0·10.

wind speeds of 5·7 and 6·7 m/sec respectively are some 50 per cent. above the values indicated by the line in Figure 9, for a sea more nearly in equilibrium with the wind.

(*b*) *Comparison of the Drag Coefficients with Published Values*

The values of drag coefficient for neutral stability from the Bass Strait observations (Fig. 5) may now be compared with the results of previous work. Charnock (1951) has given a diagram conveniently summarizing the state of knowledge at that time and Francis (1954) has collected together the results of some more recent studies.

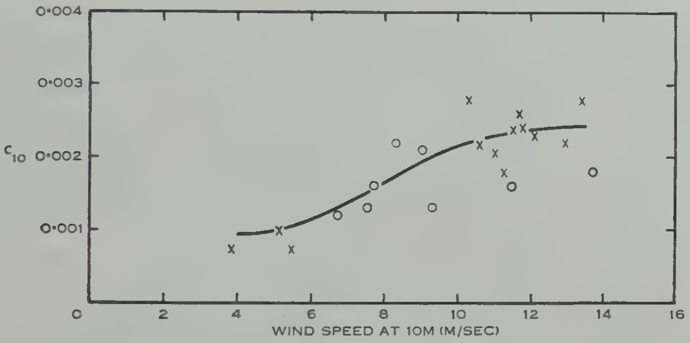


Fig. 9.—Drag coefficient related to wind speed; near-neutral observations; $-0\cdot02<Ri(13:4)<0\cdot03$.

○ Bay observations; × Bass Strait observations.

Observation of the wind-induced surface tilt of lakes or arms of the sea has given in the wind speed range 10–15 m/sec (for which our information in Figure 5 is most definite) values of c_7 scattered between 0·0015 and 0·0033 but averaging about 0·0025 as compared with $c_{10}=0\cdot0021$ given by the present work after allowing for motion of the vessel. The difference is less than 20 per cent. when the difference in wind speed reference level (rather uncertain in the tilt observations) is taken into account. At lower wind speeds we may doubt with Francis (1954) whether the tilt method is capable of giving results of much value owing to the smallness of the slopes, the interference of surges, and probably also the lack of homogeneity of the water.

A preliminary study (Sheppard and Omar 1952) by the geostrophic departure method using pilot balloon data for the Trades has given a mean value of c of about 0·0013 over the range 4–10 m/sec. This is a little lower than indicated by Figure 5 but not by a significant amount in view of the large observational scatter.

The evidence from profile studies is rather conflicting. Roll (1949) finds cogent reasons to suppose that inadequate exposure of anemometers has vitiated the results of some studies and Brocks (1955) shows, as is also evident from the present work, that more attention must be paid to atmospheric stability conditions than has been customary in much of the earlier work. This is particularly important in the light wind and large height range.

Rossby and Montgomery (1936) inferred that in light winds the sea behaves as a hydrodynamically smooth surface, $c \sim 0 \cdot 0009$, while at higher wind speeds the value of c_{10} is about $0 \cdot 0029$. Sverdrup (1946) put forward evidence from humidity gradient observations over the sea suggesting that the corresponding increase in evaporation coefficient occurs rather abruptly at about 6–7 m/sec and Munk (1947) attributed this to a marked change at that speed in the nature of the sea surface associated with the appearance of white caps. With the accumulation of further evidence it now seems doubtful if the change is as sudden as Sverdrup and Munk then supposed. Brocks (1955) in a careful study of the available data on water vapour profiles finds that the weight of evidence is in favour of the sea surface behaving as hydrodynamically smooth in light winds but that with increasing wind the evaporation and drag coefficients increase to higher values rather gradually—a conclusion in accordance with the present observations. Some wind profile observations by Hay (1955) give a similar indication; he finds that c_{10} rises from about $0 \cdot 0015$ at $6 \cdot 5$ m/sec to $0 \cdot 0023$ at 11 m/sec, a variation paralleling fairly closely that of our curve in Figure 5. Hay's observations were, however, made with a short fetch (800 m) of the wind over the sea and to leeward of steep cliffs.

Observations by Cox and Munk (1954) on the mean square slope of the water surface measured from aerial photographs of the Sun's glitter on the sea, suggest a gradual rather than a sudden increase in c with wind speed. We find, however, that the data for slope in the direction of the wind indicate a rate of increase of mean square slope with wind speed which is significantly greater above 7 or 8 m/sec (at $12 \cdot 5$ m height) than below. Munk (1955), assuming a linear variation between slope and wind speed over the whole range, has inferred that the drag coefficient probably increases with wind speed, but rather less rapidly than linearly.

Little is to be found in the literature on the variation of drag coefficient with stability. Tilt observations on Lough Neagh (Darbyshire and Darbyshire 1955) suggest that the effect is quite large, but this may be due to the use of wind observations at a nearby land station instead of over the Lough. So the variation displayed in Figure 5 remains to be more fully investigated.

VII. THE OBSERVATIONS OF VERTICAL TEMPERATURE GRADIENT

The relationship between the vertical temperature gradient in the air and the difference in temperature between sea and air is of interest, not least in order to be able to estimate the vertical temperature gradient, which is not an easily measurable quantity, from simple observations of air and sea temperatures. Furthermore, much information on the air-sea temperature difference is available from ships' routine observations; to be able to interpret these in terms of the temperature gradient in the air is of value in studies of heat transfer between sea and atmosphere, smoke diffusion, radio wave propagation, and so on.

The observations of potential temperature difference ($\Delta\theta$) between $12 \cdot 6$ and 4 m are plotted against the air-sea temperature difference in Figure 10;

various ranges of 10 m wind speed are distinguished, as a unique relationship between the temperature differences is not to be expected. For wind speeds above 5·5 m/sec any variation of the dependence is too small to be detected in the presence of considerable observational scatter and the line in Figure 10 is drawn to represent this range. This is sufficiently closely approximated by the following equations :

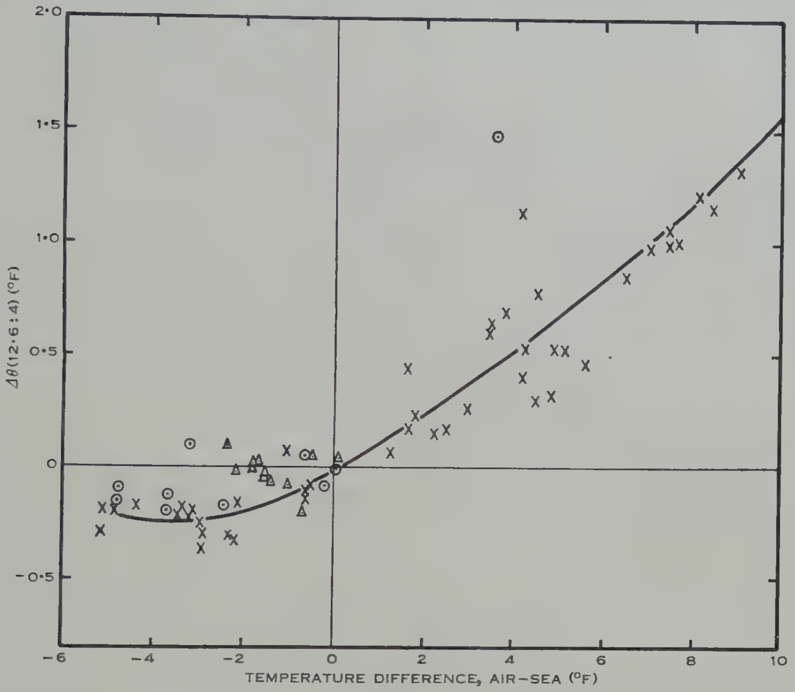


Fig. 10.—Potential temperature difference between 12·6 and 4 m in relation to the difference in temperature between sea and air.

× $u_{10}>5\cdot5$ m/sec ; ⊙ u_{10} 4·5 to 5·5 m/sec ; △ u_{10} 2·5 to 4·5 m/sec.
The line is drawn to represent the variation for $u_{10}>5\cdot5$ m/sec.

$$\left.\begin{aligned}\Delta\theta(12\cdot6:4)&=0\cdot005(T_a-T_s)^2+0\cdot105(T_a-T_s)\,; &\quad 0&<(T_a-T_s)<10\ ^\circ\mathrm{F},\\ \Delta\theta(12\cdot6:4)&=0\cdot02(T_a-T_s)^2+0\cdot14(T_a-T_s)\,; &\quad -5&<(T_a-T_s)<\ 0\ ^\circ\mathrm{F},\end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ (9)$$

where T_a=air temperature at about 3 m height,
T_s=sea surface temperature,

and temperatures are in degrees Fahrenheit. These relationships were used to estimate values of $\Delta\theta$ for some of the runs for which measured values were not available.

For the sea warmer than the air and wind speed less than 5·5 m/sec the evidence in Figure 10 is for a progressive decrease of $\Delta\theta$ such that in the lighest wind range, 2·5–4·5 m/sec, $\Delta\theta$ is not significantly different from zero. A change in this sense is to be expected from the well-established variation in the form of profiles from the logarithmic form at high wind speeds to forms at lower speeds in which a greater part of the temperature difference occurs near the boundary surface. But the magnitude of the effect is rather surprising and some doubt might be felt as to the adequacy of

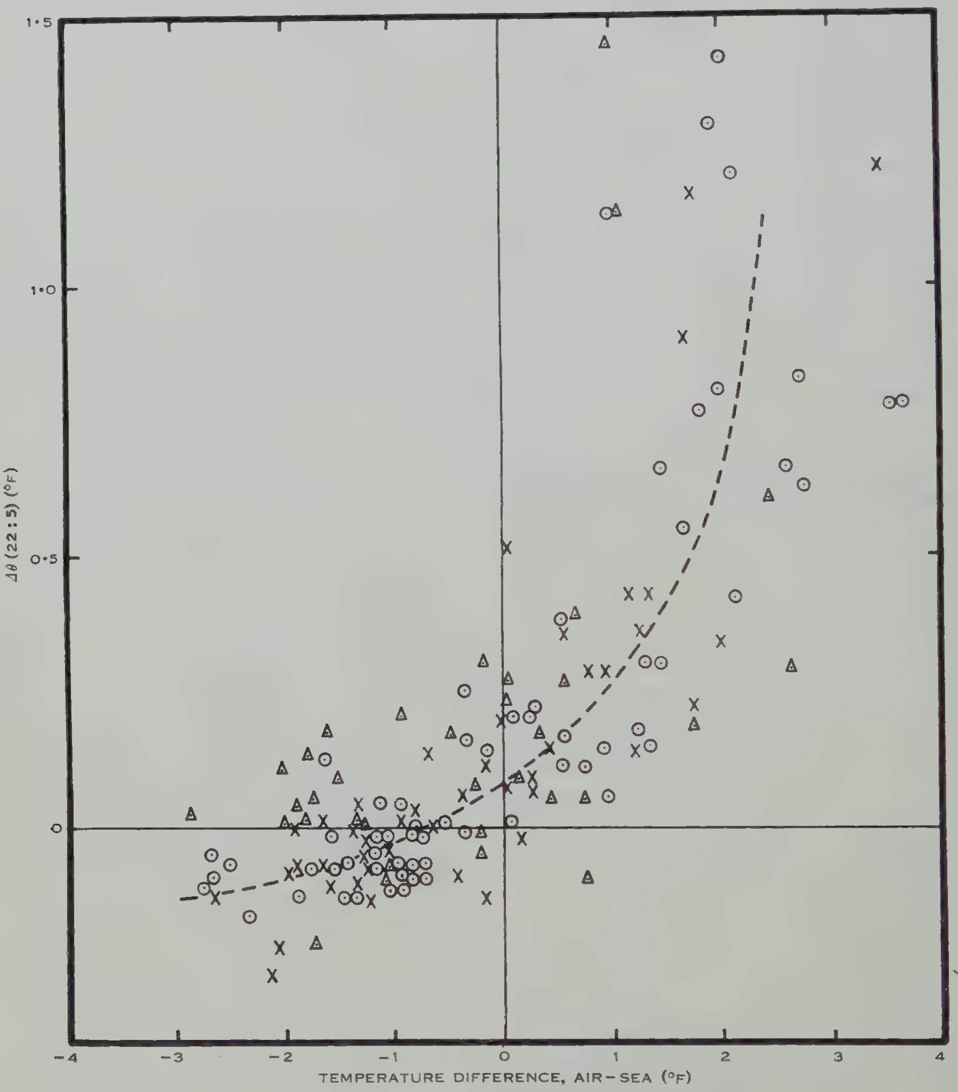


Fig. 11.—The potential temperature difference between 22 and 5 m in relation to the difference in temperature between sea and air (after Johnson and Meredith).
△ Wind speed 0–2 m/sec ; ⊙ 3–5 m/sec ; × above 5 m/sec.

ventilation of the temperature elements at low wind speeds were it not for the fact that observations by Johnson and Meredith (unpublished data 1927) on a ship in the Mediterranean show the same behaviour. They used aspirated and shielded platinum resistance elements as described by Johnson (1927) and they were of the opinion that their observations of the temperature difference 22 m – 5 m were correct to within 0·1 °F. Their counterpart of Figure 10 is reproduced here as Figure 11 and it will be seen that, with light winds and sea warmer than the air, the values of $\Delta\theta(22:5)$ are often zero or even somewhat positive. That this is by no means impossible is now becoming evident. Priestley (1954) in a study of free convection has shown how a mechanism of convective

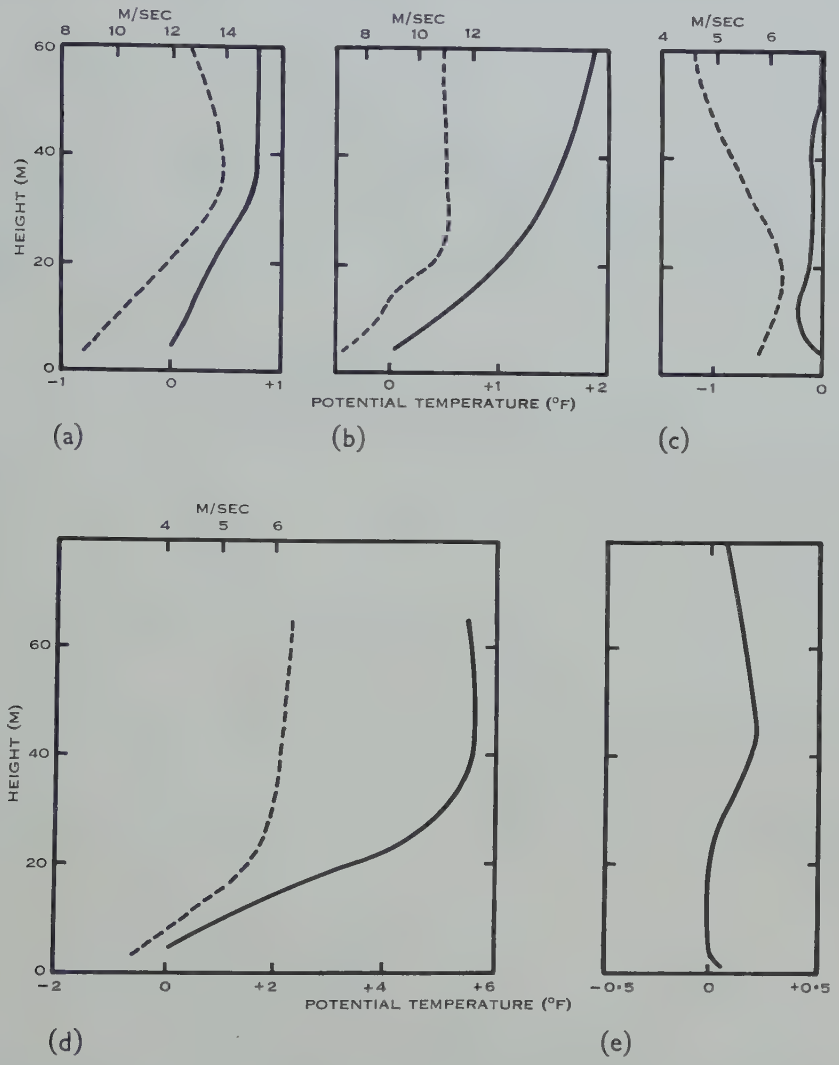


Fig. 12.—Profiles of wind speed (broken lines) and potential temperature, referred to 4 m level (full lines).

(*a*) 24.x.55, 10·50–13·00 E.S.T., Bass Strait, $T_a - T_s = 4$ °F ;
(*b*) 24.x.55, 13·30–15·40 E.S.T., Bass Strait, $T_a - T_s = 5$ °F ;
(*c*) 26.x.55, 08·40–10·30 E.S.T., Bass Strait, $T_a - T_s = -5$ °F ;
(*d*) 23.x.55, 14·20–15·20 E.S.T., Port Phillip Bay, $T_a - T_s = 3{\cdot}3$ °F ;
(*e*) 13.x.55, 12·30–15·40 E.S.T., Port Phillip Bay, $T_a - T_s = -1{\cdot}7$ °F.

plumes is able to carry heat upwards through a layer of subadiabatic lapse rate and Bunker (1956) has observed upward heat fluxes through stably stratified air at heights from 100 to 500 m over the sea.

VIII. The Kite Balloon Observations

Some wind and temperature profiles using the kite balloon equipment are shown in Figure 12; these are for the occasions when the most complete sets of data were secured. It had been hoped to obtain many more observations but troubles with balloon and cable were only surmounted a few days before the end of the trials. Each profile in Figure 12 is the mean of three sets of readings, i.e. three separate soundings, but even so the time spent by the instrument at any one level was rather small and sampling errors were probably rather large.

The temperature profiles show no very remarkable features but the tendency for the wind profiles to show maxima at low levels is unexpected. The low level maximum on October 26 (Fig. 12 (*c*)) under unstable conditions is, however, not significant as only two of the three profiles obtained on this occasion show it. The wind observations in Port Phillip Bay at the time of the temperature profiles of Figure 12 (*e*) were rather remarkable; the 13 m anemometer gave wind speeds mainly between 2 and 4 m/sec yet at heights from 30 to 110 m no wind speed of more than 1 m/sec was observed. The wind profiles taken on the ship on this day have accordingly not been analysed for resistance coefficient.

On some occasions of strong stability in Port Phillip Bay there was evidence of marked changes of wind direction with height. In particular on October 17, around 15·00 E.S.T., when the 13 m wind speed was about 10 m/sec and the air 8–9 °F warmer than the water, a streamer on the balloon cable showed that at 70 m the wind direction was about 70–80° veered from that shown by a vane at 13 m on the cross-trees. Unfortunately the balloon cable parted before more observations could be secured on this interesting phenomenon.

IX. Conclusions

The main conclusions drawn from this work may be summarized as follows:

(*a*) It is possible to obtain satisfactory observations of wind profiles on a ship despite some disturbance to the wind flow caused by the ship's hull if measurements are made both with the ship nearly stationary and also steaming into wind at moderate speeds. The interference effect may then be eliminated as shown in Section IV (*a*). It is also advisable to take records of ship motion, particularly roll, so that corrections may be made (see Section IV (*b*)).

(*b*) The tentative indication from the wind profiles is that the drag coefficient of the open sea under neutral conditions of atmospheric stability is close to 0·001 at low wind speeds but from about 5 m/sec upward to 12 m/sec there is a fairly gradual increase in resistance coefficient to a value of about 0·0021 (anemometer height 10 m). Further work is clearly needed to consolidate these findings.

(*c*) The need for atmospheric stability to be taken fully into account in all work on turbulent transfer between sea and atmosphere is evident from this work.

(*d*) The temperature gradient measurements show that with little wind and sea warmer than atmosphere the layer of appreciably superadiabatic lapse rate is very shallow, i.e. 10 m or less.

X. ACKNOWLEDGMENTS

Thanks are due to the Division of Fisheries and Oceanography, C.S.I.R.O., for making the Fishery Research Vessel *Derwent Hunter* available for this work despite a crowded programme and to the master, Capt. R. Downie, and crew for their ready cooperation.

Mr. N. G. Richards and Mr. J. Stevenson assisted on the trials and in the analysis of results.

P. A. Sheppard acknowledges a grant from the Royal Society Nuffield Foundation which allowed cooperation on this work. He is also appreciative of the hospitality of the Commonwealth Scientific and Industrial Research Organization to the same end.

XI. REFERENCES

BROCKS, K. (1955).—*Arch. Met., Wien* A **8**: 354.
BUNKER, A. F. (1956).—*Aust. J. Phys.* **9**: 133.
CHARNOCK, H. (1951).—*Sci. Progr.* **39**: 80.
COX, C., and MUNK, W. (1954).—*J. Opt. Soc. Amer.* **44**: 838.
CRAWFORD, S. G. (1951).—*J. Sci. Instrum.* **28**: 36.
DARBYSHIRE, J., and DARBYSHIRE, MOLLIE (1955).—*Quart. J.R. Met. Soc.* **81**: 333.
DEACON, E. L. (1953).—*Geophys. Mem., Lond.* **11**, No. 91.
DEACON, E. L. (1955).—C.S.I.R.O. Aust. Div. Met. Phys. Tech. Pap. No. 4.
FRANCIS, J. R. D. (1954).—*Quart. J.R. Met. Soc.* **80**: 438.
HAY, J. S. (1955).—*Quart. J.R. Met. Soc.* **81**: 307.
JOHNSON, N. K. (1927).—*Quart. J.R. Met. Soc.* **53**: 59.
MUNK, W. H. (1947).—*J. Mar. Res.* **6**: 203.
MUNK, W. H. (1955).—*Quart. J.R. Met. Soc.* **81**: 320.
PASQUILL, F. (1949).—*Proc. Roy. Soc.* A **198**: 116.
PRIESTLEY, C. H. B. (1954).—*Aust. J. Phys.* **7**: 202.
PRIESTLEY, C. H. B. (1955).—*Quart. J.R. Met. Soc.* **81**: 139.
RIDER, N. E. (1954).—*Phil. Trans.* A **246**: 481.
ROLL, H. U. (1949).—*Ann. Met., Hamburg* **2**: 71.
ROSSBY, C. G., and MONTGOMERY, R. B. (1935).—*Pap. Phys. Oceanogr.* **3**, No. 3.
ROSSBY, C. G., and MONTGOMERY, R. B. (1936).—*Pap. Phys. Oceanogr.* **4**, No. 3.
SHEPPARD, P. A., and OMAR, M. H. (1952).—*Quart. J.R. Met. Soc.* **79**: 303.
SVERDRUP, H. U. (1936).—*Geofys. Publ.* **11**, No. 7.
SVERDRUP, H. U. (1946).—*J. Met.* **3**: 1.
VAN DORN, W. (1953).—*J. Mar. Res.* **12**: 249.

APPENDIX I

TABLE 4

GENERAL CONDITIONS OF OBSERVATIONS

Date (Oct. 1955)	Location	Run Nos.	Times (E.S.T.)	Duration of Runs (min)	Clouds	Swell (ft)	Heights of Anemometers (m)			
							a	b	c	d
6	Bay	1–4	14·05–15·26	10	As 8/8		3·0	1·5	—	—
7	Bay	5–13	12·51–16·10	10	Cu 1/8, Ac As 5/8–8/8		3·0	2·67*	—	—
8	Strait	14–18	11·15–14·46	10 or 15	Cu Sc 6/8–8/8	3–4	3·0	2·67†	1·85	—
9	Bay	19–25	15·14–17·16	10	Cu Sc 7/8		4·0	2·67†	1·5	1·5‡
10	Strait	26, 27	12·08–12·52	10	Cu Sc 3/8	4–5	4·15	4·10‡	—	—
10	Strait	28–31	13·20–16·05	30	Cu Sc 6/8–7/8	4–5	4·15	4·10‡	—	—
11	Bay	32–38	11·30–16·21	30	Cu Sc Ac 4/8–7/8		4·0‡	2·64†	1·5‡	—
12	Bay	39–45	11·06–15·50	30	Cu Cb Sc 4/8–7/8		4·0‡	3·95†	1·95†	—
13	Bay	46–55	12·15–15·52	10	Cu Sc 1/8–3/8		5·81†	3·95†	1·95†	—
14	Bay	56–60	10·56–14·36	30	Ac As 8/8		4·15	—	—	—
15	Strait	61–66	11·37–15·39	30	Cu Cb Sc 4/8–7/8	4	4·15	—	—	—
17	Bay	67–75	11·41–15·12	10	Ci 7/8		4·15	—	—	—
18	Strait	76–81	12·05–15·39	10 or 30	Cu 1/8	5	4·19	2·78†	—	—
20	Bay	82–89	11·27–16·16	30	Cb As 8/8		4·19	3·68†	—	—
21	Bay	90	13·20–13·50	30	Cu Sc 6/8		5·85†	3·68†	—	—
22	Bay	91–100	10·30–16·05	30	Cu 1/8–4/8		5·85†	4·23	3·68†	—
23	Bay	101–104	12·56–14·27	30	Cu 1/8, Ac 1/8		5·85†	4·23	3·68†	—
24	Strait	105–112	10·48–15·47	30	Cu Sc 1/8–2/8	8	5·85†	4·23	3·68†	—
26	Strait	113–116	08·04–10·51	30	Cu Cb 5/8–7/8	8	5·85†	4·23	—	—

---

* This anemometer was rigged forward of the anemometer mast by 4·2 m in runs 5–7, by 2·7 m in runs 8–10, and by 1·5 m in runs 11–13.

† Rigged 2·7 m forward of anemometer mast using boom of $1\frac{1}{2}$ in. pipe.

‡ On these occasions the arms on the anemometer mast were extended in length from the original 0·6 m to 1·5 m and the longer arms were used for the anemometer at heights between 4·15 and 4·23 m on all occasions after October 13.

TABLE 5

WIND AND TEMPERATURE OBSERVATIONS AND COMPUTED RESISTANCE COEFFICIENTS*

Run No.	Wind Direction (deg. T.B.)	Fetch (naut. mile)	Wave Height (ft)	T_a—T_s (°F)	Δθ (°F)	$u(13)$ (cm/sec)	$u(6\cdot4)$ (cm/sec)	$u(a)$ (cm/sec)	$u(b)$ (cm/sec)	$u(c)$ (cm/sec)	$u(d)$ (cm/sec)	Δu (cm/sec)	$Ri(13:4)$ $\times 10^3$	$u(10)$ (m/sec)	c_{10} $\times 10^4$
1	005	22(1)	—	3·3	(0·5)	609	547	506	455			90	(100)	5·9	14
2	005	22(1)	1·5	4·2	(0·6)	770	691	629	569			117	(75)	7·4	17
3	355	20(1)	1·5	4·6	(0·6)	784	641	628	561			128	(60)	7·5	22
4	340	20(1)	1·5	4·1	(0·6)	770	672	604	547			136	(55)	7·4	26
5	345	12	2	2·9	—	527	439	376	382			123	—	5·0	—
6	345/360	12	2	3·4	—	576	488	420	425			128	—	5·5	—
7	355	13	2·5	4·0	(0·6)	644	564	498	499			123	(67)	6·1	30
8	350	14	2·5	4·7	0·82†	834	731	655	650			145	53	8·0	26
9	350	11	2·5	4·6	1·16†	830	726	647	641			149	71	7·9	25
10	350	11	2·5	4·5	1·73†	904	792	706	705			166	86	8·6	24
11	000	14	4	3·8	1·07†	899	812	742	712			133	81	8·7	16
12	350	14	3–4	4·0	0·99†	966	856	771	748			163	51	9·3	25
13	000	13	3	4·8	0·77†	950	855	777	752			146	49	9·2	20
14	220	Large	1	—0·15	—0·10†	524	498	470	456			47	0	5·1	10
15	220	,,	1	+0·1	0·05†	384	377	347	344	340		30	0	3·8	7
16	220	,,	1·5	0·0	—0·02†	553	530	504	491	486		42	0	5·4	7
17*a*	220	,,	1·5	—0·6	—0·12†	700	671	626	615	607		56	—70	6·9	20
17*b*	220	,,	2	—0·6	—0·17†	668	640	598	586	579		52	—110	6·6	22
18*a*	210	,,	2	—0·6	—0·10†	753	731	664	672	655		59	—50	7·4	17
18*b*	210	,,	2	—0·6	—0·07†	731	711	647	646	637		59	—40	7·2	15
18*c*	210	,,	2	—0·6	—0·12†	720	694	630	627	—		67	—40	7·1	23
19	200	10(3)	1·5	—2·2	(—0·25)	655	631	610	583	562	570	47	(—200)	6·5	23
20	200	9(1½)	1·5	—2·5	(—0·25)	656	628	608	578	560	570	51	(—170)	6·5	26
21	200	10(2½)	1·5	—2·3	—0·31	694	661	638	611	584	590	57	—160	6·8	29
22	200	10(2½)	1·5	—2·4	(—0·25)	619	600	—	552	531	542	41	(—250)	6·1	21
23	200	8(1)	1·5	—1·9	(—0·25)	647	620	596	569	546	552	51	(—160)	6·4	28

TABLE 5 (*Continued*)

Run No.	Wind Direction (deg. T.B.)	Fetch (naut. mile)	Wave Height (ft)	T_a—T_s (°F)	Δθ (°F)	$u(13)$ (cm/sec)	$u(6\cdot4)$ (cm/sec)	$u(a)$ (cm/sec)	$u(b)$ (cm/sec)	$u(c)$ (cm/sec)	$u(d)$ (cm/sec)	Δu (cm/sec)	$Ri(13:4)$ $\times 10^3$	$u(10)$ (m/sec)	c_{10} $\times 10^4$
24	200	8(½)	1·5	—2·2	—0·32	730	699	672	647	615	628	56	—180	7·2	27
25	200	6(¼)	1·5	—3·0	—0·37	699	666	646	614	586	590	57	—200	6·9	31
26	200	Large	4	—2·2	(—0·24)	656	613	—	—			79	(—60)	6·4	—
27	200	,,	4	—1·9	(—0·24)	702	656	—	—			85	(—60)	6·9	—
28	200	,,	—	—2·0	(—0·23)	656	611	574	570			85	(—60)	6·4	—
29*a*	200	,,	—	—3·5	(—0·25)	680	648	609	608			68	(—90)	6·7	—
29*b*	200	,,	4	—3·4	(—0·25)	694	652	609	607			84	(—60)	6·8	—
30	210	,,	—	—3·4	(—0·25)	751	712	665	665			83	(—60)	7·4	—
31	210	,,	—	—3·4	(—0·25)	766	717	668	669			96	(—50)	7·5	—
32	210/225	8(4½)	1	—4·8	—0·10	499	480	453	460	438		34	—150	4·9	19
33	205/220	8(4½)	1	—4·8	—0·16	465	451	422	431	412		28	—350	4·6	20
34	200/220	9(5)	1	—3·6	—0·13	484	465	438	444	427		34	—190	4·8	22
35	200/210	9(4½)	1·25	—3·2	0·10	538	515	484	488	468		40	—160	5·3	24
36	195/200	9(5)	1·5	—3·1	—0·20	672	648	610	609	586		49	—140	6·6	21
37	195/205	8(4½)	1·5	—3·2	—0·21	611	588	552	555	528		44	—180	6·0	23
38(*s*)	195	8(4½)	1·25	—2·4	—0·17	537	521	483	484	474		36	—220	5·3	21
39	200/220	11(4)	2·5	—4·4	—0·18	868	835	788	786	752		72	—60	8·5	20
40	200/205	10(4)	2·5	—3·4	—0·21	823	795	750	748	717		65	—85	8·1	21
41	200/210	10(3)	2·5	—3·4	(—0·17)	959	917	860	862	817		87	(—40)	9·4	20
42	210	10(2½)	2·5	—2·2	—0·16	1002	955	898	898	848		95	—30	9·8	18
43	205	9(2)	2·5	—3·3	—0·18	968	923	866	866	826		90	—40	9·5	21
44	200	9(2½)	2·5	—3·0	—0·25	1049	996	936	939	889		99	—45	10·3	23
45	210	9(2½)	2·5–3·0	—2·9	—0·30	1112	1069	1007	1003	943		100	—50	10·9	23
46	270	12(3)	1	—1·5	—0·01	345	337	331	319	290		—	—	2·6	—
47	270	10(1¾)	1	—1·5	—0·05	262	260	253	248	218		—	—	4·0	—
48	250	6(1½)	0·75	—1·4	—0·06	227	226	217	211	181		—	—	2·3	—
49	260	6(½)	0·5	—1·7	0·37	98	93	84	86	39		—	—	0·9	—

TABLE 5 (*Continued*)

Run No.	Wind Direction (deg. T.B.)	Fetch (naut. mile)	Wave Height (ft)	$T_a—T_s$ (°F)	Δθ (°F)	$u(13)$ (cm/sec)	$u(6\cdot4)$ (cm/sec)	$u(a)$ (cm/sec)	$u(b)$ (cm/sec)	$u(c)$ (cm/sec)	$u(d)$ (cm/sec)	Δu (cm/sec)	$Ri(13:4)$ $\times 10^3$	$u(10)$ (m/sec)	c_{10} $\times 10^4$
51	025/030	20	0·5	—2·1	—0·02	354	355	344	334	310		—	—	3·5	—
52	040	19	0·5	—2·3	0·10	316	319	308	298	275		—	—	3·2	—
53	035	18	0·5	—1·6	0·03	337	335	330	319	296		—	—	3·4	—
54	035	18	0·75	—1·8	0·03	402	397	388	376	354		—	—	4·0	—
55	035	18	0·75	—1·8	0·01	392	386	374	362	350		—	—	3·9	—
56	000	27	4	1·6	0·45	934	855	807				131	15	9·0	21
57	000	26	5	2·5	0·18	1178	1100	1037				142	15	11·5	16
58	000	24	5	3·5	0·61	1271	1182	1117				156	42	12·4	14
59	000	24	6	4·2	0·40	1412	1300	1230				187	19	13·7	18
60	000	24	5	—	0·07	1055	962	908				152	37	10·2	19
61	270	14	5–6	2·0	—0·04	1388	1259	1180				215	(4)	13·4	28
62	290	13	4–6	1·9	(0·10)	1208	1099	1038				178	(5)	11·7	26
63	270	10	6	1·0	0·05	1335	1230	1156				180	2	13·0	22
64	260	12	5	1·6	—0·19	1160	1075	1015				147	(8)	11·3	18
65	295	11	5	2·5	—0·10	1133	1039	983				156	(14)	11·0	21
66	290	8	3	3·7	(0·30)	1095	999	938				160	(20)	10·6	22
67	000	20(1)	2	6·5	(0·85)	1131	1001	936	—			204	(34)	10·8	31
68	000	20(½)	1·5–2	6·5	0·85	1109	1004	949	—			168	51	10·7	20
69	000	20(1½)	2	7·7	1·02	1073	961	900	—			179	54	10·3	24
70	000	20(1½)	2	7·5	1·00	989	882	827	—			169	59	9·5	24
71	000	20(1)	2	7·0	0·99	912	805	752	—			167	60	8·7	28
72	000	20(1)	2	7·6	1·07	908	803	744	—			168	64	8·7	28
73	000	19(½)	2	8·5	1·17	1281	—	1069	—			217	42	12·3	26
74	010	20(1½)	2	8·2	1·22	1091	—	899	—			197	53	10·4	28
75	005	20(1)	2	9·0	1·33	1235	—	—	—			—	—	11·8	—
76	230	Large	5	—1·2	—	318	310	295	—			—	—	3·2	—
77	230	,,	5	—0·7	—0·20	265	253	241	—			—	—	2·6	—

TABLE 5 (*Continued*)

Run No.	Wind Direction (deg. T.B.)	Fetch (naut. mile)	Wave Height (ft)	T_a—T_s (°F)	$\Delta\theta$ (°F)	$u(13)$ (cm/sec)	$u(6\cdot4)$ (cm/sec)	$u(a)$ (cm/sec)	$u(b)$ (cm/sec)	$u(c)$ (cm/sec)	$u(d)$ (cm/sec)	Δu (cm/sec)	$Ri(13:4)$ $\times 10^3$	$u(10)$ (m/sec)	c_{10} $\times 10^4$
78	210/190	Large	5	—1·0	—0·08	341	348	334	—			—	—	3·5	—
79	190/020	,,	4–6	—1·0	0·07	638	613	582	534			—	—	6·3	—
80	190/220	,,	4–6	—0·6	0·06	545	533	510	461			—	—	5·4	—
81	215	,,	4–6	—0·5	0·06	332	332	319	284			—	—	4·5	—
82	350	24	3–4	4·6	0·77	827	758	711	694			121	89	8·0	15
83	340/010	22	3–4	3·8	0·69	757	673	636	614			131	68	7·2	24
84	350/300	21·5/ 12(4)	3–4	3·5	0·65	1077	985	917	902			163	41	10·4	21
85	290	10(3½)	3–4	0·2	0·04	1192	1138	1064	1052			—	—	11·7	—
86	310/330	9(3½)	3–4	—0·7	—0·01	586	536	527	497			76	—3	5·7	21
87	360/060	16	2	—0·8	0·04	685	627	614	589			85	9	6·6	17
88	040/180	16/5	2	0·5	0·07	370	331	314	287			61	32	3·6	26
89	—	—	—	—	(0·03)	332	293	287	259			60	(25)	3·2	34
90	310	19	3	2·6	0·01	857	792	740	733			118	6	8·3	22
91	300	12(½)	3–4	—1·2	(—0·18)	967	921	—	882	870		87	(—40)	9·5	20
92	300	12(½)	3–4	—0·6	—0·14	1015	966	—	918	910		96	—26	10·0	14
93	300	12(5½)	3–4	—0·1	—0·06	951	899	887	863	847		94	—11	9·3	13
94	300	12(4¾)	3–4	—	(0)	794	744	740	718	707		81	(0)	7·8	16
95	300	11(4)	3–4	0·9	0·01	774	722	722	688	686		85	12	7·6	13
96	300	11(5)	3	1·3	0·02	692	644	642	615	614		76	20	6·8	12
97	290/280	10(3)	3	1·7	0·04	728	672	671	642	640		86	39	7·1	13
98	270/290	11(1½)	2	1·8	0·05	634	569	568	540	541		96	42	6·1	21
99	280	11(2½)	2	—	0·04	630	562	570	543	537		91	40	6·1	19
100	280	10(1)	2–1·5	2·4	0·03	594	524	524	491	493		105	25	5·7	32
101	320	21(5)	4	5·7	—	767	682	671	634	—		139	—	7·3	—
102	320	21(5)	4	4·2	1·13	671	580	570	540	526		138	100	6·4	29
103	320/360	21	3	3·6	1·48	449	374	360	341	322		119	175	4·2	—

TABLE 5 (*Continued*)

Run No.	Wind Direction (deg. T.B.)	Fetch (naut. mile)	Wave Height (ft)	$T_a—T_s$ (°F)	$\Delta\theta$ (°F)	$u(13)$ (cm/sec)	$u(6{\cdot}4)$ (cm/sec)	$u(a)$ (cm/sec)	$u(b)$ (cm/sec)	$u(c)$ (cm/sec)	$u(d)$ (cm/sec)	Δu (cm/sec)	$Ri(13:4)$ $\times 10^3$	$u(10)$ (m/sec)	c_{10} $\times 10^4$
104	360	21	3	2·9	1·90	496	408	394	368	345		139	165	4·6	—
105	260	24	—	3·9	—	930	855	832	795	803		130	(30)	9·0	19
106	260	24	3–5	3·0	0·27	1251	1148	1115	1063	1073		182	14	12·1	23
107	270	15	5	4·5	0·30	1201	1084	1055	997	1010		178	16	11·6	24
108	270	14	4–5	4·8	0·32	1069	969	946	893	901		171	18	10·3	28
109	270	12	5	5·6	0·47	927	826	817	771	774		156	33	8·9	18
110	260	14	3–4	4·9	0·53	937	843	822	775	778		159	35	9·0	17
111	260/250	13/15	3	4·3	0·54	877	789	770	724	729		148	42	9·4	21
112	260	15	3–4	5·1	0·53	1200	1089	1064	1009	1009		187	26	11·6	24
113	230/260	Large/22	4	—5·2	—0·31	653	600	594	572			81	—72	6·4	—
114	240/210	18/large	—	—4·9	—0·20	713	670	667	639			72	—59	7·0	—
115	210	Large	—	—5·1	—0·18	645	600	593	576			70	—56	6·3	—
116	210	Large	—	—3·7	—0·20	515	479	472	460			57	—95	5·0	—

* *Notes to Table* 5.—In the column headed " Fetch " two values are given in those cases where the first part of the fetch of the wind was over shallow water; the value in parentheses is the fetch over deep water.

Wave heights were estimated. They are more reliable for Port Phillip Bay than in Bass Strait, where the swell made estimation of wave height more difficult.

$T_a—T_s$ is the difference in temperature between air and sea and $\Delta\theta$ is the potential temperature difference between 4 m and 12·6 m except for those values marked with a dagger where the lower level is 3 m. Values of $\Delta\theta$ in parentheses were estimated from $T_a—T_s$ using the curve of Figure 10.

$u(a)$, $u(b)$, etc. signify the corrected wind speeds relative to the sea surface at the heights a, b, etc. given in Table 4.

Δu is the difference in wind speed between 13 and 4 m as read from the smoothed wind profiles.

$Ri(13:4)$ is the layer Richardson number for the 4 to 13 m air layer calculated from the values of $\Delta\theta$ and Δu. Values in parentheses are based on estimated values of $\Delta\theta$.

c_{10} is the drag coefficient, the reference height for wind speed being 10 m.

Values of the drag coefficient have not been calculated for either October 13 (run Nos. 46–55) or for October 18 (run Nos. 76–81) owing to the combination of marked instability and a variable light wind on the former occasion and the marked decrease of wind above 13 m shown by the balloon anemometer on the latter.

Drag coefficients are also not given for October 10 (run Nos. 26–31) or for October 26 (run Nos. 113–116) owing to rather heavy rolling in winds of less than 8 m/sec—a condition giving too great an error (see Section IV (*b*)) for trustworthy results.

APPENDIX II

Method of Evaluation of the Shearing Stress from the Non-neutral Wind Profiles using Rossby's Equation

In applying Rossby's equation

$$\frac{u_*}{z\partial u/\partial z}=\frac{k}{(1+\sigma Ri(z))^{\frac{1}{4}}} \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{A1})$$

to the evaluation of u_* from measurements of wind speed u and potential temperature θ at two heights, a and b, we proceed by using

$$\partial u/\partial z=cz^{-\beta} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{A2})$$

as a suitable interpolation formula between these levels. Rearranging equation (**A1**) and integrating with respect to height between levels a and b gives

$$\frac{k^2}{u_*^2}\int_a^b z^2\left(\frac{\partial u}{\partial z}\right)^4 \mathrm{d}z=\frac{\sigma g}{T}\int_a^b \frac{\partial\theta}{\partial z}\mathrm{d}z+\int_a^b \left(\frac{\partial u}{\partial z}\right)^2 \mathrm{d}z.$$

Using equation (**A2**) and the expression derived from it that

$$c=\frac{(1-\beta)(u_b-u_a)}{b^{1-\beta}-a^{1-\beta}}$$

gives

$$u_*^2=\frac{Fk^2(u_b-u_a)^2}{G+\sigma Ri(b:a)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{A3})$$

where

$$F=\frac{(1-\beta)^4(p^{3-4\beta}-1)(p-1)}{(3-4\beta)(p^{1-\beta}-1)^4},$$

$$G=\frac{(1-\beta)^2(p^{1-2\beta}-1)(p-1)}{(1-2\beta)(p^{1-\beta}-1)^2},$$

$$p=b/a,$$

and $Ri(b:a)$ signifies the Richardson number for the layer a to b obtained by using $(u_b-u_a)/(b-a)$ for the wind velocity gradient and similarly for the potential temperature gradient. It now remains to be able to assign values to β. With very accurate wind measurements at a number of heights ranging from a to b it would be possible to evaluate β from them but this is impossible in the present work so β is calculated from equation (**A1**). To do this an assumption must now be made as to the form of the temperature profile and observation suggests

$$\partial\theta/\partial z=dz^{-\delta}$$

as a reasonable approximation. Using the fact that

$$\beta=-z\partial^2u/\partial z^2(\partial u/\partial z)^{-1}$$

and similarly for δ, differentiation of equation (**A1**) with respect to z readily gives

$$\beta=\frac{1+\sigma Ri(z)+\frac{1}{2}\delta\sigma Ri(z)}{1+2\sigma Ri(z)}. \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{A4})$$

Lacking information on δ we now have to make the further assumption of similarity of the profiles of u and θ so that $\delta=\beta$, in which case

$$\beta=\frac{1+\sigma Ri(z)}{1+1\cdot 5\sigma Ri(z)}. \qquad \text{(A5)}$$

The assumption of $\beta=\delta$ is unsatisfactory owing to the fact that there are differences in the transfer mechanisms for heat and momentum, but for the sea and over the range of stability here considered ($-0\cdot 03<Ri<0\cdot 10$) it is likely that the difference is not large and the nature of equation (**A4**) is such that small differences are not important. Taking, for example, the case of $\sigma=9$ and $Ri(z)=0\cdot 11$ we see from equation (**A4**) that values of δ ranging from $0\cdot 7$ to $0\cdot 9$ only result in a range of β from $0\cdot 78$ to $0\cdot 82$.

Values of β calculated using equation (**A5**) with values of the layer Richardson number $Ri(b:a)$ are considered to be appropriate to the layer a to b for use in equation (**A3**) and in this way values of the proportionality constant between u_*^2 and $(u_b-u_a)^2$ as a function of $Ri(b:a)$ are derived. These are given in Table 2 of Section V of the present paper for the particular case of $a=4$ m and $b=13$ m.

STATISTICAL ANALYSIS OF AUSTRALIAN TEMPERATURE DATA

By S. C. Das*

[*Manuscript received July 5, 1956*]

Summary

The mean summer maximum temperatures of five Australian inland localities for a period of 65 years from 1891 to 1955 are analysed and it is shown that the temperatures have an overall parabolic trend.

I. Introduction

In a previous paper (Das 1956) the statistical analysis of Australian pressure data was considered in order to test a hypothesis, regarding a shift in the high pressure belt, which arose out of Deacon's (1953) work. It was shown there that the belt is slowly moving southward. In this paper we attempt to make a statistical analysis of the mean summer maximum temperatures of five Australian inland localities to verify how far the temperature data conform to the above hypothesis. The five localities chosen for the purpose were Bourke, Hay, Broken Hill, Narrabri, and Walgett. Broken Hill has data dating back to 1891, while the other localities have data for a few more years. The mean summer maximum temperatures as given in the data are the averages of the mean monthly maxima for the months of December, January, and February. For example, the figure given for 1891 is the average of the mean monthly maxima for December 1890, January and February 1891.

II. Analysis of the Data

In order to examine whether the data follow a trend, a polynomial regression line of the form $y=A'+B'\xi_1'+C'\xi_2'$, or of the form $y=A'+B'\xi_1'+C'\xi_2'+D'\xi_3'$ depending on the tests of goodness of fit of temperature on time, has been fitted to the data for individual localities. Following the accepted notation used in Fisher and Yates's tables, ξ_1', ξ_2', and ξ_3' stand for orthogonal polynomials of degree one, two, and three respectively in t, where t is the number of years measured from the centre point of the series. The results are shown in Table 1.

The second, third, and fourth columns give the three regression coefficients and the next three columns give the corresponding t values for testing the significance of these regression coefficients. The first regression coefficient is significant for each locality and the second is significant only in the case of Hay. The last column gives the corresponding regression equations. It is evident from the regression equations that in all of the stations the mean summer

* Australian National University, Canberra; present address: Ravenshaw College, Cuttack, Orissa, India.

maximum temperatures exhibit slow linear decreases except for Hay where the trend is parabolic.

In testing $t_{C'}$ for Hay for significance we must, of course, regard it as having been chosen as the largest of the five statistics. Nevertheless it is clearly still significant at the 5 per cent. point even after allowing for this. With the exception of Bourke all the values of $t_{C'}$ are positive, moreover, which suggests that a stronger conclusion regarding the effect of the variable ξ_2' on the temperatures may be got by a joint test of significance of the five $t_{C'}$ values.

TABLE 1

Locality	B'	C'	D'	$t_{B'}$	$t_{C'}$	$t_{D'}$	Regression Equations (°F)
Bourke	—0·06983	—0·00003		3·58	0·09		$y=97\cdot4-0\cdot06983t$
Hay	—0·02353	0·00402	0·00003	3·60	2·94	1·28	$y=88\cdot5-0\cdot02353t+0\cdot00402t^2$
Broken Hill ..	—0·06979	0·00180		4·40	1·90		$y=90\cdot4-0\cdot06979t$
Narrabri ..	—0·03227	0·00033		3·72	1·06		$y=93\cdot7-0\cdot03227t$
Walgett ..	—0·02081	0·00151		2·38	0·78		$y=94\cdot7-0\cdot02081t$

We may consider the model

$$\begin{pmatrix} \varkappa_{1t} \\ \varkappa_{2t} \\ \vdots \\ \varkappa_{5t} \end{pmatrix} = \mathbf{B} \begin{pmatrix} \xi_{0t}' \\ \xi_{1t}' \\ \xi_{2t}' \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{5t} \end{pmatrix}$$

or

$$\boldsymbol{\varkappa}_t = \mathbf{B}\boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t,$$

where $\varkappa_{jt}$ is the temperature at the jth station at time t and $\mathbf{B}$ is a (5×3) matrix of regression coefficients. The ε_t's are presumed to be serially independent and to have a joint normal distribution with an arbitrary (non-singular) correlation matrix. We then wish to test the hypothesis that the last column of $\mathbf{B}$ is composed entirely of zeros. The likelihood ratio test of this hypothesis is the partial canonical correlation of the vector $\boldsymbol{\varkappa}_t$ with ξ_{2t}' when the effects of the other elements in $\boldsymbol{\xi}_t$ have been removed (Bartlett 1947). In this simple case the partial canonical correlation reduces to the partial multiple correlation of ξ_{2t}' with $\varkappa_{1t}, \varkappa_{2t}, \ldots, \varkappa_{5t}$ when the effects of means and ξ_{1t}' have been removed. The test statistic may be computed as

$$F=\frac{1-L}{L}\cdot\frac{n-7}{5},$$

where

$$L=\frac{|\mathbf{C}_{\varkappa\varkappa\cdot1}-[\mathbf{C}_{\varkappa2\cdot1}\mathbf{C}_{\varkappa2\cdot1}']C_{22\cdot1}^{-1}|}{|\mathbf{C}_{\varkappa\varkappa\cdot1}|},$$

and

$$\mathbf{C}_{\varkappa\varkappa\cdot 1}=\mathbf{C}_{\varkappa\varkappa}-[\mathbf{C}_{\varkappa 1}\,.\,\mathbf{C}_{1\varkappa}]C_{11}^{-1},$$

$$\mathbf{C}_{\varkappa 2}=\left[\sum_{t=1}^{65}\varkappa_{jt}\xi_{2t}'\right],\quad \text{(a matrix of 5 rows)}$$

$$\mathbf{C}_{\varkappa\varkappa}=\left[\sum_{t=1}^{65}(\varkappa_{it}-\bar{\varkappa}_i)(\varkappa_{jt}-\bar{\varkappa}_j)\right],\quad \text{(a } 5\times 5 \text{ matrix)}$$

$$\mathbf{C}_{1\varkappa}=\left[\sum_{t=1}^{65}\xi_{1t}'\varkappa_{jt}\right].\quad \text{(a matrix of 5 columns)}$$

$$C_{11}=\sum_{1}^{65}\xi_{1t}^{2},\ C_{22\cdot 1}=C_{22}=\sum_{1}^{65}\xi_{2t}^{2}.$$

In the present case

$$F=\frac{58}{5}\,\cdot\,\frac{1-L}{L}=3\cdot 70.$$

This is distributed in the F distribution with 5 and 58 degrees of freedom and is therefore highly significant. There is no doubt therefore that there is some overall second degree effect whose influence appears to be to slow down and reverse the effect of the decline in maximum summer temperatures.

III. ACKNOWLEDGMENT

The author wishes to thank Mr. E. L. Deacon, Division of Meteorological Physics, C.S.I.R.O., for suggesting the problem.

IV. REFERENCES

BARTLETT, M. S. (1947).—*J.R. Stat. Soc. Suppl.* **9**: 176–97.
DAS, S. C. (1956).—*Aust. J. Phys.* **9**: 394.
DEACON, E. L. (1953).—*Aust. J. Phys.* **6**: 209–18.

A RELATION BETWEEN SNOW COVER, CIRRUS CLOUD, AND FREEZING NUCLEI IN THE ATMOSPHERE

By E. G. Bowen*

[*Manuscript received August* 2, 1956]

Summary

It is reasonable to suppose that observations like that of cirrus cloud in the upper air and heavy falls of snow in relatively warm latitudes correspond to the presence of a large number of freezing nuclei in the atmosphere. A 300-year record of snow covering the ground at Tokyo and a 10-year record of cirrus cloud in Western Australia are examined and compared with one year's measurement of freezing nucleus concentration. The curves show a high degree of correlation, and all three tend to maximize on certain calendar dates.

These dates are nearly identical with those on which singularities occur in world rainfall. It has previously been suggested that this phenomenon is due to the nucleating effect of dust from meteor showers.

The record of snow cover at Tokyo for the three days December 29, 30, and 31 is examined separately and found to show a distinct recurrence tendency, which, over the 300 years of record, is identical in period and phase with that of the Bielid meteor shower.

I. Introduction

In a recent publication, Arakawa (1956) described how the first time that snow completely covered the ground in Tokyo was for many years an occasion for seasonal greetings and the payment of feudal dues, and that the date has therefore been recorded over a long period of time. The record is an interesting one for two reasons. It covers a period of more than 300 years, which is considerably longer than the majority of meteorological records. Secondly, among the many meteorological requirements which must be satisfied for snow to cover the ground at Tokyo, it is probably essential that a large number of freezing nuclei should exist in the atmosphere.

Interest in the role played by freezing nuclei in the atmosphere has lain dormant for many years but was revived by the work of Schaefer (1946) and Cwilong (1947), and more recently by the work of Ôkita (1952) and Isono (1955) on the chemical composition of some of the particles responsible. Interest has again been aroused by the suggestion that particles of meteoritic dust might act as freezing nuclei when they fall into the lower atmosphere and in this way have an influence on the weather at the Earth's surface (Bowen 1956*a*). This proposal has proved unacceptable to the majority of meteorologists and a number of papers have appeared in the literature arguing that it is incorrect. (See, for example, Martyn (1954), Neumann (1954), Swinbank (1954), Oliver and Oliver (1955), and Bhalotra (1956).) The difficulty which many meteor-

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Chippendale, N.S.W.

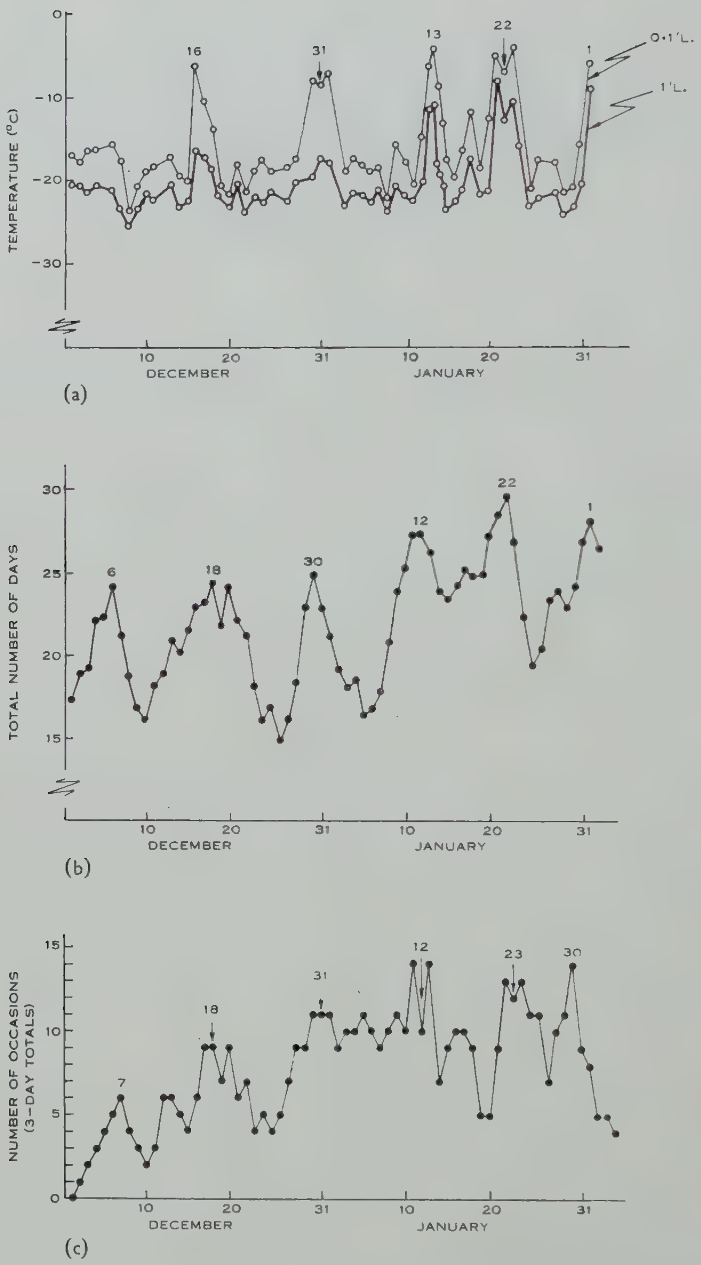


Fig. 1 (*a*).—Concentration of freezing nuclei measured on the ground in Western Australia, December 1955 and January 1956.

Fig. 1 (*b*).—Occurrence of cirrus cloud at six stations in Western Australia over a 10-year period.

Fig. 1 (*c*).—Date on which snow first covered the ground at Tokyo, from 1632 to 1955.

ologists find in accepting the hypothesis is no doubt partly due to a lack of appreciation of the part played by freezing nuclei in atmospheric processes, and partly to the absence of a long series of freezing nucleus measurements upon which a judgment might be based.

If the meteoritic dust hypothesis is correct, it would be expected that:

(*a*) high freezing nucleus counts would tend to occur almost simultaneously in different parts of the world,

(*b*) they would repeat on approximately the same calendar dates each year.

Measurements of freezing nucleus concentration made in several parts of the world in January 1954, 1955, and 1956 tend to confirm these expectations (Bowen 1956*b*), but a longer series of observations is probably required before this is generally accepted.

One purpose of the present paper is to point out that certain of the more conventional meteorological observations may serve as indicators of the presence of large numbers of freezing nuclei in the atmosphere. Among these are records such as that of snow cover in Tokyo, which has already been referred to, and the occurrence of cirrus cloud in the atmosphere.

II. Cirrus Cloud in Western Australia

The existence of cirrus clouds indicates that, among other things, there must at the time of observation be a considerable number of freezing nuclei in the upper air. In Figure 1 (*a*) is a curve of freezing nucleus concentration measured by Bigg (1956) in a cold chamber on the ground at Carnarvon in Western Australia during the months of December 1955 and January 1956. It shows the temperature at which 1 and 0·1 freezing nucleus per litre respectively were observed in that locality. In Figure 1 (*b*) is a curve of the total number of days on which cirrus cloud was observed at six stations distributed over the western half of Australia during a 10-year period. It is obvious by inspection that there is a high degree of correlation between them, and Bigg has shown this to be so quantitatively. This suggests that cirrus cloud is a good indicator of a high freezing nucleus content in the atmosphere, and that over a 10-year period the count has tended to be high on certain calendar dates.

In the more densely populated areas of the world, low cloud frequently interferes with the observation of cirrus cloud. It is unlikely, therefore, that long unbroken records of this kind will be found except in arid regions like Arizona, the Sahara, and the western part of Australia.

III. First Snow Cover in Tokyo

As already mentioned, the record of snow cover at Tokyo might also indicate days on which there was a high freezing nucleus concentration in that part of the world. This particular record refers to a different hemisphere from that in which the measurements just discussed were made, to a different season of the year, and to an entirely different climatological situation. According to accepted meteorological concepts, therefore, it would not be expected that these results would bear any relation to those in Western Australia. On the other hand, if meteoritic dust were involved, one might expect a relationship between them.

In Figure 1 (*c*) are the 3-day totals of the date of the first appearance of snow cover at Tokyo for the period 1632 to 1955, taken from Tables 1 and 2 of Arakawa's paper. As there is only one observation per season, as compared with one per day in Figure 1 (*b*), the statistical fluctuations are considerable, but in Figure 1 (*c*) there is a distinct preference for certain dates and the major peaks appear to correspond closely with those in the previous two curves. Cross correlation of curves (*b*) and (*c*) gives a correlation coefficient of 0·33, which is significant at the 99 per cent. level. This suggests, with a high degree of probability, that the association between the curves is a real one and that there is some physical phenomenon which is common to both. The relation which each curve bears to Figure 1 (*a*) indicates that this common factor may be the concentration of freezing nuclei in the atmosphere.

It is also apparent from Figure 1 that both cirrus cloud in Western Australia and snow cover in Japan tend to maximize on certain calendar dates during December and January; and that these dates are almost identical in the two regions, which are in different hemispheres, in entirely different geographical situations, and subject to entirely different climatic regimes. It is difficult to explain this behaviour except through the agency of an extraterrestrial phenomenon such as meteoritic dust. Finally, the fact that the curves in Figure 1 refer to 1, 10, and 323 years of observation respectively suggests that the effect has been present for several centuries.

IV. RELATION TO RAINFALL SINGULARITIES

In the first column of Table 1 are the dates in Figure 1 (*b*) on which cirrus cloud is a maximum. In the second column are the dates on which the major peaks in the Tokyo snowfall record of Figure 1 (*c*) are centred. There are smaller peaks on December 13 and January 5 and 16 which may or may not have any significance and have been neglected. It has already been pointed out (Bowen 1956*b*) that during these same months there are singularities in

TABLE 1

PEAK DATES IN CIRRUS CLOUD, SNOW COVER, AND WORLD RAINFALL

Date of Maximum in Cirrus Cloud in Western Australia	Date of Maximum in Snow Cover at Tokyo	Date of Maximum in World Rainfall
December 6	December 7	December 5
18	18	14
30	31	December 29 January 1
January 12	January 12	January 12
22	23	22
February 1	January 30	February 1

world rainfall on the dates given in the third column of Table 1. It will be observed that they are nearly identical in all three cases. If the three phenomena were random and independent of each other, the probability of an association as close as this occurring by chance may be calculated to be less than one in a million.

It has also been shown (Bowen 1956*b*) that the peaks in world rainfall occur approximately 30 days after prominent meteor streams. It is not proposed to pursue this relationship in the present paper, but to point out a special property of one of these meteor streams, namely the Bielids, and the precipitation thought to be associated with it.

V. Periodicity Associated with the Bielid Meteor Stream

The Bielid meteor stream is outstanding among those in the latter part of the year for its very pronounced periodic behaviour. It is associated with Biela's comet which is known to have a 6·6-year period of rotation around the Sun and the meteors have made many strong appearances at 6- or 7-year intervals. A list of the years in which either the comet or the meteor stream has been seen is given in the first column of Table 2 (Porter 1952). Where there are closely spaced appearances, as in 1845, 1846, and 1847, the mean is taken.

Table 2

BIELA'S COMET AND METEORS, AND SNOW COVER AT TOKYO

Date of Appearance of Biela's Comet and Meteors	Snow Covering the Ground at Tokyo on December 29, 30, and 31
	1672
	1731
1741	1742
	1756
1772	
	1778
	1785
	1792
1798	1798
1805	
1830 1832 } 1831	
1838	
1845 1846 } 1846 1847	
1852	
	1859
1867	
1872	
1885	1884
1899	
	1905

There are many gaps in the sequence, which may have been due to failure of the meteors to appear, to unsuitable observing conditions, or to other reasons. Nevertheless a distinct recurrence tendency at 6- to 7-year intervals is apparent.

There does not appear to be a generally recognized method of serially correlating such data and the following procedure has therefore been adopted. The years given in the first column of Table 2 are marked on two horizontal time scales and successively displaced 0, 1, 2, 3, 4, . . ., etc. years in the usual way. The coincidence of two events is arbitrarily given a score of 2, and events separated by one year a score of 1. All other separations score zero. Proceeding in this way gives the correlogram of Figure 2 (*a*) which shows a very definite recurrence tendency in the data, with a period between 6 and 7 years.

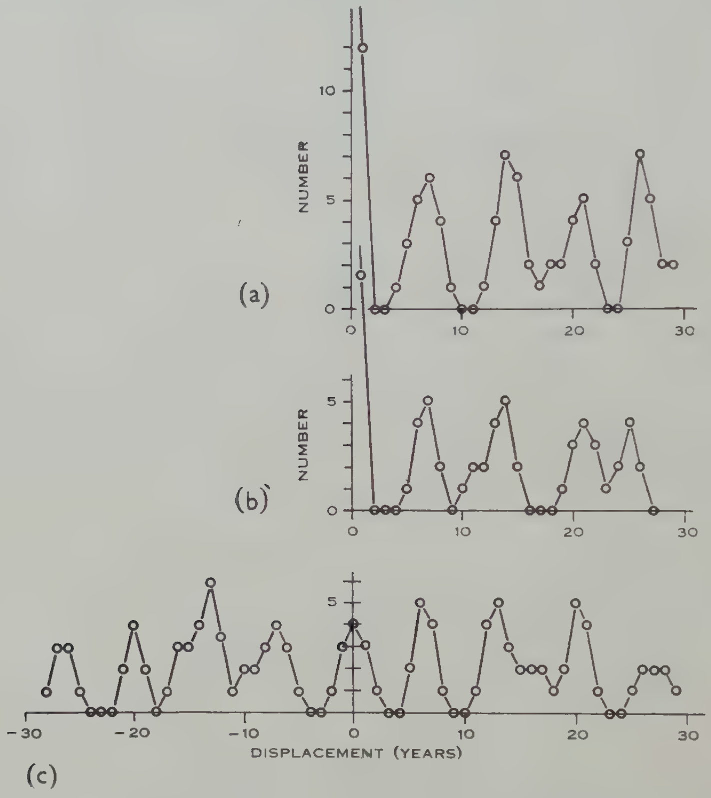


Fig. 2 (*a*).—Serial correlation of the years on which Biela's comet or associated meteors appeared.

Fig. 2 (*b*).—Serial correlation of years on which snow first covered the ground at Tokyo on any of the three days December 29, 30, and 31.

Fig. 2 (*c*).—Cross correlation of snow cover at Tokyo on December 29, 30, and 31, and Biela's comet and meteors.

It has been shown (Bowen 1956*a*) that there is a similar recurrence tendency in the rainfall of several parts of the world during the three days December 29, 30, and 31, which are approximately 30 days after this meteor shower. An investigation will now be made of the snow cover data for Tokyo to see if the same phenomenon exists.

In the second column of Table 2 are the years on which the first snow cover occurred at Tokyo during the same three days, December 29, 30, and 31. Again there are breaks in the sequence, due to the many conditions of temperature, synoptic situation, etc., all of which have to be satisfied before a heavy fall of snow is recorded. If, now, the dates are serially correlated in the same way as the comet and meteor data, the correlogram of Figure 2 (*b*) is obtained, which again shows a distinct recurrence tendency with a period between 6 and 7 years.

Finally, if the years of snow cover and those on which meteors appeared are cross-correlated, the correlogram given in Figure 2 (*c*) is obtained. It shows the same periodic behaviour with a significance at least as great as that in Figure 2 (*a*) and its symmetry about zero displacement indicates that the phase of the two occurrences is identical.

VI. CONCLUSIONS

It appears therefore that:

(i) Records of cirrus cloud and the date on which snow first covers the ground in Tokyo are good indicators of the existence of high freezing nucleus counts in the atmosphere.

(ii) During December and January there is a close relationship between cirrus cloud in Western Australia, snow cover in Japan, and world rainfall, which is difficult to explain except on the basis of an extraterrestrial influence acting on all three quantities.

(iii) The incidence of snow cover at Tokyo on December 29, 30, and 31 shows, over a period of 300 years, a recurrence tendency which is identical in period and phase with that of the Bielid meteor stream.

VII. REFERENCES

ARAKAWA, H. (1956).—*Quart. J.R. Met. Soc.* **82**: 222–6.

BHALOTRA, Y. P. R. (1956).—*Weather* **11**: 48–51.

BIGG, E. K. (1956).—New evidence of a meteoritic influence on rainfall. *Tellus* (in press).

BOWEN, E. G. (1956*a*).—*J. Met.* **13**: 142–51.

BOWEN, E. G. (1956*b*).—*Nature* **177**: 1121–3.

CWILONG, B. M. (1947).—*Proc. Roy. Soc.* A **190**: 137–43.

ISONO, K. (1955).—*J. Met.* **12**: 456–62.

MARTYN, D. F. (1954).—*Aust. J. Phys.* **7**: 358–64.

NEUMANN, J. (1954).—*Aust. J. Phys.* **7**: 522–6.

ÔKITA, T. (1952).—*Sci. Rep. Tôhoku Univ. Ser. 5, Geophysics* **4** (2): 78–82.

OLIVER, M. B., and OLIVER, V. J. (1955).—*Bull. Amer. Met. Soc.* **36**: 147–51.

PORTER, J. G. (1952).—"Comets and Meteor Streams." (Chapman and Hall: London.)

SCHAEFER, V. J. (1946).—*Science* **104**: 457–9.

SWINBANK, W. C. (1954).—*Aust. J. Phys.* **7**: 354–8.

JANUARY FREEZING NUCLEUS MEASUREMENTS

By E. G. Bowen*

[*Manuscript received June* 29, 1956]

Summary

Measurements of the concentration of freezing nuclei in the atmosphere have been made in widely separated parts of the globe during January in one or more of the three successive years 1954, 1955, and 1956. The results of all three sets of observations are summarized and it is shown that high freezing nucleus counts tend to occur on dates which are identical or nearly identical with those on which unusually heavy falls of rain have previously been shown to occur. The results are discussed in relation to the possible connexion between meteoritic dust and world rainfall.

I. Introduction

The suggestion has been made that dust from meteor streams when it falls into the cloud systems of the lower atmosphere acts as a nucleating agent and induces unusually heavy amounts of rain to fall. Examination of the world rainfall records for the first 5 weeks of the year suggests that the dates on which heavy rainfall occurs are approximately January 12 and 22 and February 1. If the hypothesis is correct, it would be expected that the freezing nucleus concentration in the atmosphere would also be high on or about these dates.

Measurements of freezing nucleus concentration were made in Sydney in January 1954 and again in Sydney and in other parts of the world in January 1955. The results have been reported by Cwilong (1955) and by Smith, Kassander, and Twomey (1956) and they gave sufficient support to the hypothesis to encourage further measurements during January 1956. These were carried out in South Africa, on the west and east coasts of Australia, and on the west coast of the U.S.A. The results are reported in the present group of four papers.

The South African measurements were undertaken at Pretoria by Dr. S. C. Mossop of the South African National Physical Laboratory and those in the United States of America by Dr. R. N. Bracewell of Stanford University. The Australian measurements were carried out by members of the staff of the Radiophysics Laboratory at Carnarvon on the north-west coast and at Sydney on the east coast. The detailed results of the 1956 measurements are given in the three papers which follow. The purpose of the present paper is to summarize the results obtained in the whole series of observations made during 1954, 1955, and 1956 and to comment on the method of measurement.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Chippendale, N.S.W.

II. Scope of the Measurements

Freezing nuclei may be defined as particles in the atmosphere on which ice crystals can grow from the vapour phase or, alternatively, as materials which, in solution in a water droplet or in contact with it, will induce it to freeze. They become effective at some temperature below 0 °C, the actual value of which depends on their chemical composition and the immediate past history of the surface exposed to the atmosphere. Schaefer (1950) has shown that a number of the materials commonly found on the Earth's surface are active at temperatures about −15 to −20 °C, while certain other materials, of which silver iodide is an example, are effective at comparatively warm temperatures of about −4 °C.

The method of measuring the number of freezing nuclei in the atmosphere and the temperature at which they become effective is to reproduce in a cold chamber the process which usually occurs in a supercooled cloud. A sample of air is taken, it is saturated with water vapour, and the temperature is reduced. A cloud of water droplets forms and, after a short interval, a comparatively small number of ice crystals appears. An estimate is then made of the number of freezing nuclei per litre and the temperature at which they become active.

It is likely that freezing nucleus measurements may be adversely affected by terrestrial dust and contamination from industrial sources. In order to be representative of what happens in clouds, measurements are preferably made in aircraft above the haze layer. When aircraft are not available, a reasonable alternative is to make measurements at ground level in air of recent maritime origin. The 1956 measurements in Pretoria and Sydney were made in aircraft at heights up to 20,000 ft; those in Western Australia were made on the ground in air which had had a long travel over the Indian Ocean; and those at Palo Alto were also made on the ground in air which was predominantly from the Pacific Ocean.

The South African and Sydney measurements were made in aircraft of the South African Air Force and the Royal Australian Air Force respectively, in a cold chamber specially designed for aircraft operation by Smith and Heffernan (1954). The measurements in Western Australia and on the west coast of the United States of America were made by a new method devised by Bigg and described in one of the accompanying papers. This method had the advantage of being an objective one, it was more sensitive than the methods which had previously been employed, and was particularly valuable in that it was capable of measuring small concentrations of very active particles.

There were systematic differences between the mean readings taken in aircraft and on the ground. These appear to have been due partly to the different geographical locations but more particularly to the different methods of measurement employed. In the aircraft equipment the temperature of the cold chamber fell continuously at a rate of 4 deg/min, so that a crystal which formed at any given temperature would not be seen and counted until a colder temperature had been reached. In Bigg's cold chamber, however, the temperature was maintained at one value during the whole period an ice crystal was growing. These differences bring out the importance of using a standardized method of measurement in any future series of observations, in order to determine whether there are any systematic differences in the freezing nucleus concentration in one part of the world as compared with another.

III. Summary of Results

An outstanding characteristic of measured values of freezing nucleus concentration is that they remain at comparatively low values for up to a week or ten days at a time and then suddenly increase by orders of magnitude for one or two days. The interesting periods are those in which the freezing nucleus counts were high, and the dates on which this occurred in the present series of measurements are summarized in Table 1. In the first part of the table are the results for 1954 and 1955 which have already been given by Cwilong and Smith, Kassander, and Twomey. In the latter part of the table are the results for 1956 described in the accompanying papers. In the latter series of measurements, peaks in the freezing nucleus counts were obtained on the dates indicated in South Africa, Western Australia, and the west coast of the United States of America. At Sydney no peaks were obtained comparable in magnitude with those in 1954 and 1955. This is thought to have been due to one of two causes: either the particles responsible for high nucleus counts were absent in this area in 1956, or they may have precipitated out during the heavy rain which fell in Sydney on the few days immediately preceding the dates on which high freezing counts were observed elsewhere.

Table 1

DATES OF HIGH FREEZING NUCLEUS COUNTS IN JANUARY 1954–1956

Year	Place	Date (January)		
1954	Sydney	12, 14	22	31
1955	Sydney	13	23	29
	Hawaii	10	23	29
	Arizona	—	—	—
	Panama	13	18	31
1956	South Africa	13	—	29
	Western Australia	13	22	Feb. 1
	Sydney	—	—	—
	Palo Alto, U.S.A.	13	21	—

It is apparent from Table 1 that the peaks do not always appear in all localities, but when they do occur they are on nearly the same dates in widely separated parts of the globe. A histogram giving the 3-day totals of the dates on which the freezing nucleus peaks were observed is given in Figure 1, showing peaks on January 13, 22, and 30.*

---

* It should be observed that the majority of the measurements in this series did not commence until January 2 or 3, so that this diagram contains no information about the first three days in January.

This behaviour is that to be expected from the pattern of rainfall for the month of January for many parts of the world (Bowen 1956). In that paper it was shown that singularities in rainfall tend to occur on approximately the same dates in Australia, New Zealand, South Africa, United States of America, Great Britain, Japan, and the Netherlands. Totalling the available rainfall data from all these regions gives the curve of Figure 2, which shows peaks on January 12 and 22 and February 1, in close agreement with the freezing nucleus curve of Figure 1.

The fact that high freezing nucleus counts are obtained on identical or nearly identical dates in places separated up to 180° in longitude points to an extraterrestrial origin of the particles. The fact that the phenomenon has tended to repeat on each of the three years of observation is consistent with the meteor hypothesis.

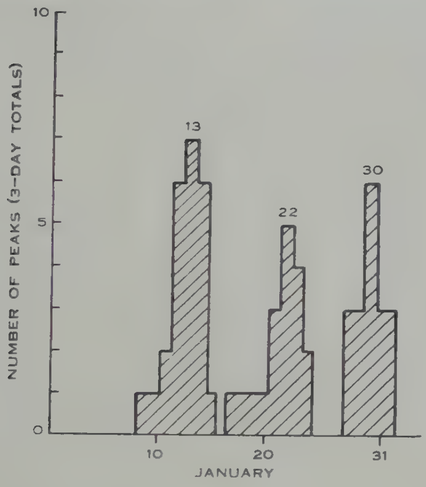


Fig. 1

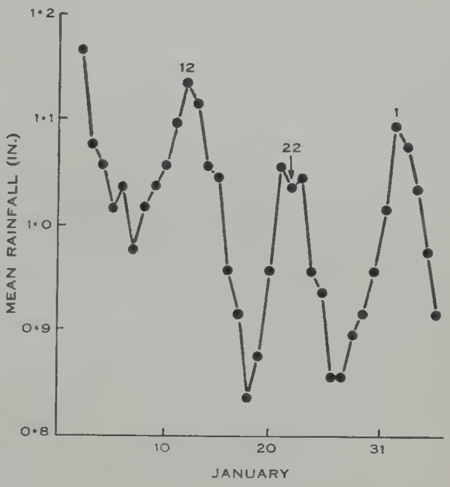


Fig. 2

Fig. 1.—Histogram of the dates in January on which high freezing nucleus counts were obtained during 1954, 1955, and 1956.

Fig. 2.—The mean daily rainfall of January for approximately 300 stations distributed over the globe.

It is thought that the present results make some contribution to the understanding of the processes in the atmosphere leading to the formation of rain, and provide encouragement to obtain further measurements of freezing nucleus concentration in many different parts of the world.

IV. References

Bowen, E. G. (1956).—*J. Met.* **13** : 142–51.
Cwilong, B. M. (1955).—*Nature* **176** : 129–30.
Smith, E. J., and Heffernan, K. J. (1954).—*Quart. J. R. Met. Soc.* **80** : 182–97.
Smith, E. J., Kassander, A. R., and Twomey, S. (1956).—*Nature* **177** : 82–3.
Schaefer, V. J. (1950).—*Z. angew. Math. Phys.* **1** : 153–84, 217–36.

COUNTS OF ATMOSPHERIC FREEZING NUCLEI AT PRETORIA, JANUARY 1956

By S. C. Mossop,* A. E. Carte,* and K. J. Heffernan†

[*Manuscript received June* 29, 1956]

Summary

An account is given of freezing nucleus measurements made in aircraft at Pretoria, South Africa, during the month of January 1956. The measurements showed variations from day to day, which are discussed in relation to the prevailing weather situation. The possible connexion with particles from meteor showers is touched upon.

I. Introduction

The present paper describes daily measurements of the concentration of freezing nuclei in the atmosphere near Pretoria over the period January 3 to February 3, 1956. The apparatus used consisted of an airborne cloud chamber of 100 l. capacity and 10 l. illuminated volume, as described by Smith and Heffernan (1954). Samples of outside air were introduced by flushing with a strong air stream for 3 min. The chamber was then cooled at a rate of 4 °C/min and steam was introduced at the bottom to maintain a cloud of uniform and reproducible density. The illuminated cloud was examined visually and the number of ice crystals which appeared was recorded as a function of temperature. Complete glaciation of the cloud was observed at about -40 °C and the chamber was then warmed up to $+10$ °C ready for the next sample of outside air to be introduced.

The flights lasted from approximately 0900 to 1130 hr and were made early in the day to avoid turbulent mixing conditions as far as possible. The usual procedure was to take three air samples at 20,000 ft, two at 10,300 ft, and finally one at 5300 ft for analysis on the ground. (The heights given are heights above sea-level, Pretoria being at 4500 ft.) All sampling was done within a 30-mile radius of Pretoria. The 0 °C isotherm was generally at about 16,000 ft.

II. Experimental Results

In Figure 1 are plotted the temperatures at which the concentration of ice crystals reached $0 \cdot 1$, 1, and 10^3 per litre. These are the average values for the samples taken at a particular height.

The $0 \cdot 1$/l line indicates the temperature at which there was, on the average, always one ice crystal in the 10-l. illuminated volume. This ignores the isolated single crystals which sometimes appear from about -5 °C downwards. Since these isolated crystals may be of great meteorological importance, the total

---

* South African National Physical Laboratory, C.S.I.R., Pretoria, South Africa.

† Division of Radiophysics, C.S.I.R.O., University Grounds, Chippendale, N.S.W.

number appearing at and above −20 and −30 °C were also plotted, the readings taken at 20,000 ft each day being averaged. These two curves are very similar in general features to the 0·1/l and 1/l curves shown in Figure 1 (*a*), so they have not been included.

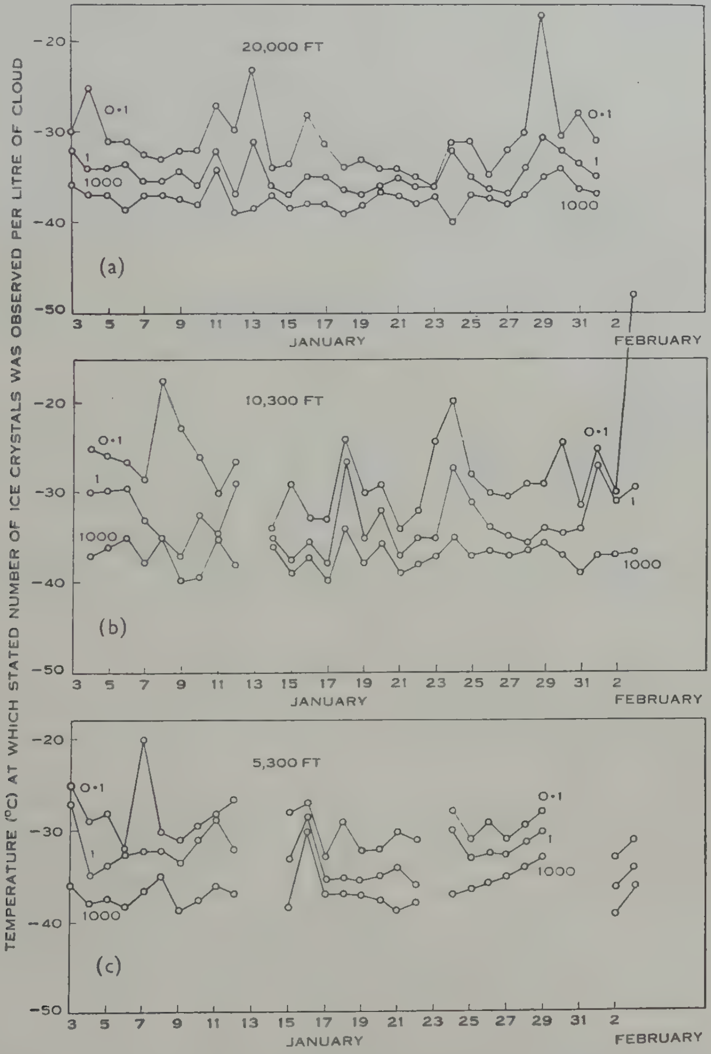


Fig. 1.—Daily plots of the temperature at which the concentration of atmospheric ice-forming nuclei reached the values indicated, at heights of 20,000, 10,300, and 5,300 ft above sea-level at Pretoria, over the period January 3 to February 3, 1956.

III. Meteorological Conditions

Since it is possible that the day-to-day fluctuations in the freezing nucleus content of the atmosphere may be considerably influenced by the previous history of the air mass in which measurements are made, the meteorological situations during the period under consideration have been examined and the nature of the air mass at 20,000 ft is described below. It must be emphasized that, owing to the paucity of upper air soundings available, the air mass analysis can seldom be made with complete certainty.

January 3.—Dry hot continental air from west.

January 4–5.—Extremely dry superior air from south, with cool moist sea air below 10,000 ft.

January 6.—Deep moist unstable air of maritime origin. Widespread showers.

January 7–12.—Scattered light showers on 7th. The maritime air was displaced by tropical continental air following an anticyclonic trajectory from the south-east. Subsidence was indicated on the 8, 9, 10, 12th by drying of the air above 10,000 ft. Column became warm and moist on 11th.

January 13–16.—Deep warm moist air, uniform to 20,000 ft, probably maritime tropical air warmed over land. Widespread heavy rain on 13th, progressively less on 14th and 15th.

January 17–27.—Extremely dry subsided upper air from south-east. On 24th and 25th, isolated thunderclouds indicated increased penetration of surface air to high levels. There was a temporary increase of moisture content on the 26th.

January 28–29.—Maritime tropical air converged over Transvaal causing the air to be moist and unstable throughout. Widespread showers on both days.

January 30–February 2.—Reversion to subsidence conditions with very dry air at 20,000 ft.

February 3.—Similar conditions but there was sufficient convective activity in lower moist layers for widespread thunderstorms to build up.

IV. Discussion of Results

There was no marked scatter in the observations at a particular level on any one day, the temperatures for particular concentrations of ice crystals agreeing with the average value to about ±2 °C. The only exception was on January 13, when the scatter among the six sets of readings at 20,000 ft for concentrations of ice crystals up to 10^2/l was about ±6 °C.

The average temperature at which the ice crystal concentration reached 1/l was −34 °C at 20,000 ft, −32·5 °C at 10,300 ft, and −33 °C at 5300 ft (500 ft above ground level at the aerodrome). Considering the day-to-day variations between individual readings it is doubtful whether the differences between these values are significant. Counts on 10 days scattered throughout December and February gave the same averages as those made during January. They are significantly lower than those made with similar apparatus at Sydney and at Arizona (Smith, Kassander, and Twomey 1956).

The meteorological conditions during January were unusual in that for long periods dry subsided air persisted above 10,000 ft and typical thunderstorm conditions were virtually absent. These periods, namely, January 7–10, January 17–27, and January 30–February 2, coincided with low counts of freezing nuclei at 20,000 ft (possibly excepting January 30 and 31). Counts were, however, just as low on six days in February when the weather situation had changed completely, and humid tropical air gave widespread rain.

There were three occasions in January when there was deep moist unstable maritime air extending above 20,000 ft, namely January 6, 13–15, and 28–29. Rain fell at more than half of 49 stations within a 40-mile radius of Pretoria during these three periods. The heaviest and most widespread rains occurred on the 13th when the mean for all stations was 15 mm. At no other times during the period January 3 to February 3 was there anything but light isolated showers.

Of these dates, the 13th and 29th coincide with peaks in the ice crystal count. If we attempt to explain these peaks as being due to freezing nuclei carried up from the ground, it is difficult to understand :

(i) the absence of high counts on the 6th,

(ii) the non-persistence of high counts to the 14th and 15th,

(iii) the fact that the count at 20,000 ft on the 29th is higher than at 10,000 ft or near the ground. (Unfortunately the lower readings had to be omitted on the 13th owing to the lack of dry ice, so the same comparison cannot be made.)

If Bowen's meteoritic dust hypothesis is correct, then the absence of high ice crystal counts on or about January 22 points to non-uniform distribution of the dust particles. Some support for this is given by the observations of January 13 when three of the six air samples gave much higher ice crystal counts than the others. In these three cases the number of particles visible in the chamber before a cloud was formed was also much higher than usual.

V. Counts of Airborne Particles

It may be of interest to mention some auxiliary experiments in which daily counts were made of the number of particles collected upon spiders' threads exposed in the air stream at 20,000 ft for about ½ hr. The average number of particles (virtually all non-hygroscopic) collected per ½ in. length of thread, corrected to the same exposure time and air-speed, showed minor peaks on January 13 and February 1 and very large increases on January 21, 23, and 24.

VI. Conclusion

To summarize, one may therefore say that significant peaks were found in the counts of atmospheric freezing nuclei at 20,000 ft on January 13 and 29 and in the counts of non-hygroscopic particles on or about January 13, 23, and February 1.

Our results cannot be fully explained either by Bowen's meteoritic dust hypothesis or by suggesting that the counts are influenced by air mass changes. It is apparent that a full explanation can only be obtained by making further surveys of this kind in various parts of the globe.

VII. Acknowledgments

The authors' thanks are due to the Australian C.S.I.R.O. for providing the apparatus with which the above experiments were carried out, and for enabling one of us (K.J.H.) to take part. The authors are deeply indebted to the South African Air Force for placing an aircraft at their disposal, and for the ready cooperation of the personnel of 28 Squadron. They thank the Director and staff of the Weather Bureau, Pretoria, for help with meteorological information and Mr. J. Taljaard in particular for making the analysis of air masses. Mr. G. J. Viljoen is thanked for help with the experimental observations.

This paper is published with the approval of the South African Council for Scientific and Industrial Research.

VIII. References

Smith, E. J., and Heffernan, K. J. (1954).—*Quart. J.R. Met. Soc.* **80** : 182–97.
Smith, E. J., Kassander, A. R., and Twomey, S. (1956).—*Nature* **177** : 82–3.

COUNTS OF ATMOSPHERIC FREEZING NUCLEI AT CARNARVON, WESTERN AUSTRALIA, JANUARY 1956

By E. K. Bigg*

[*Manuscript received June 29, 1956*]

Summary

A new instrument for measuring freezing nucleus concentration is described, and an account given of a series of measurements made at Carnarvon in Western Australia during January 1956. The measurements were made on the ground in marine air with a long travel over the Indian Ocean and are thought to be free from contamination by terrestrial dust and smoke from industrial sources.

The measurements showed distinct maxima on dates which are identical or nearly identical with those expected from Bowen's meteor theory.

I. Introduction

Freezing nucleus measurements were made at ground level in Carnarvon, Western Australia, over the same period as those in South Africa described in the preceding paper. This particular site was chosen because the weather at this time of year was known to be remarkably uniform, being practically cloudless and with persistent on-shore winds. This ensured that the air sampled would be as free as possible from contamination by smoke or terrestrial dust.

II. Method of Measurement

The observations were made by a new technique in which ice crystals formed in a cold chamber of simple design were allowed to fall into a slightly supercooled solution of cane sugar and water, where they grew large enough to be counted easily. This avoided the difficulty of making a visual estimate of the concentration of ice crystals in a beam of light as used in the South African and other experiments.

A sketch of the apparatus used in the present measurements is given in Figure 1. It consisted of a hollow cylinder surrounded by two jackets containing ethylene glycol and water mixtures. The outer of these was cooled with dry ice, and the temperature of the mixture in the inner jacket—which could be stirred vigorously to ensure uniform temperature distribution—represented the minimum temperature of the inner cylinder and hence of the cloud which it contained. The cloud chamber was kept in an insulated box until a measurement was to be made. A fresh sample of air was obtained merely by lifting it from this box, and a cloud was formed when it was placed in the position shown in the sketch, on a tank bearing a dish of supercooled sugar solution. Ice crystals generated in the cloud then fell into the solution at the base of the chamber, and after a fixed period of 2 min could easily be counted.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Chippendale, N.S.W.

Some precautions were essential. Perhaps the most important was the removal of frost from all surfaces in contact with the cloud. This was ensured by careful and frequent wiping with glycerol. Spontaneous nucleation in the sugar solution, which could occur if the cold sugar solution absorbed water from the air, or if the temperature fell too low, was avoided by frequent replacement of the sugar on humid days, and by ensuring good thermal regulation. Accurate counting was limited to a maximum of about 100 crystals with the usual method of operation, though it could be extended by slowing down the rate of crystal growth.

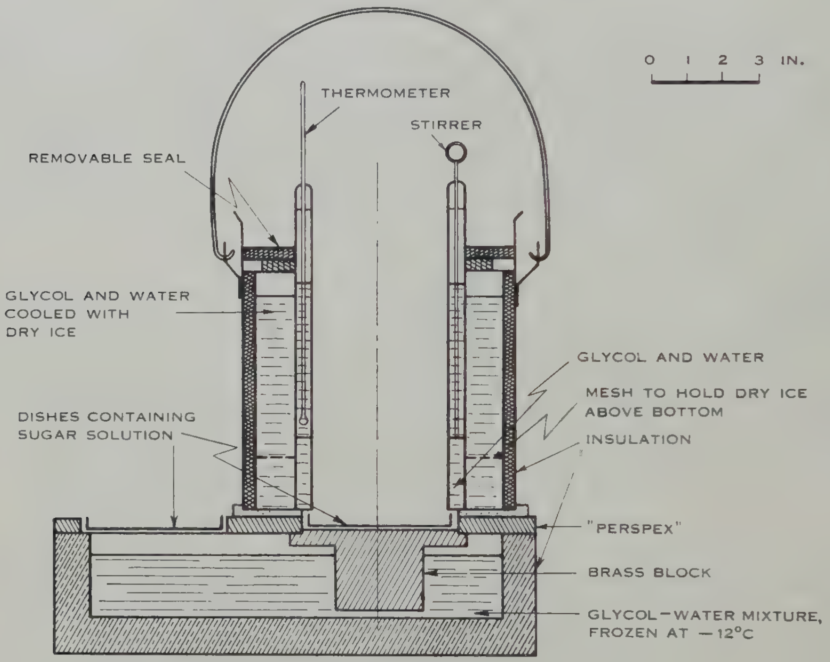

0 1 2 3 IN.
THERMOMETER
STIRRER
REMOVABLE SEAL
GLYCOL AND WATER COOLED WITH DRY ICE
GLYCOL AND WATER
MESH TO HOLD DRY ICE ABOVE BOTTOM
DISHES CONTAINING SUGAR SOLUTION
INSULATION
"PERSPEX"
BRASS BLOCK
GLYCOL-WATER MIXTURE, FROZEN AT −12°C

Fig. 1.—Cold chamber and cooling box.

Observations were made at a series of fixed temperatures and a plot of ice crystal concentration against temperature obtained. Typical measurements showed that the number of ice crystals usually increased logarithmically with decreasing temperature. Concentrations less than 1/l were estimated by performing the experiment many times at approximately constant temperature and finding an average. The temperature for a concentration of 0·1/l was deduced by this means, usually with the help of some extrapolation. The high probability of obtaining a count other than the true mean when the numbers were small was reduced by taking many observations and, except when the number of crystals did not increase logarithmically with decreasing temperature, the error from this source was probably less than ±2, ±1, ±1 °C for 0·1/l, 1/l, and 10/l respectively. On the remaining occasions it may have been twice as great.

III. Results

The temperatures at which the concentrations of ice crystals reached on the average 0·1/l, 1/l, and 10/l are plotted in Figure 2. This figure differs from Figure 1 of the preceding paper in giving integrated rather than instantaneous concentrations. A second difference is that a concentration of 10/l is shown instead of 1000/l, as counts above 100 were not usually made. Omission of the 0·1/l readings for January 14 (third experiment) and January 24 were made necessary by unusually wide scatter in the readings, while there was no dry ice on January 2 and 27.

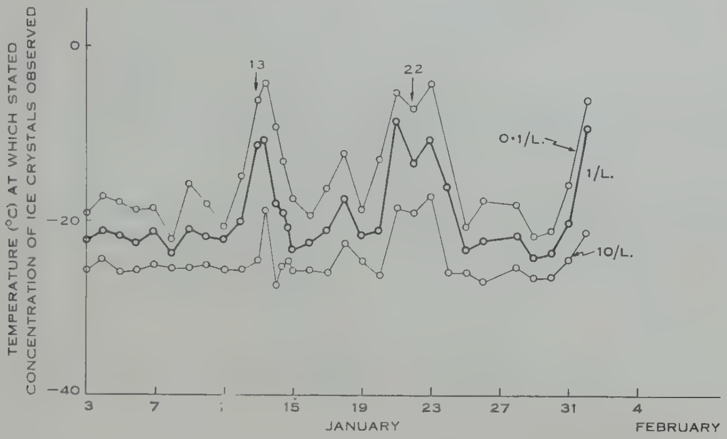


Fig. 2.—The temperatures for 0·1, 1, 10 crystals/l during January at Carnarvon, Western Australia.

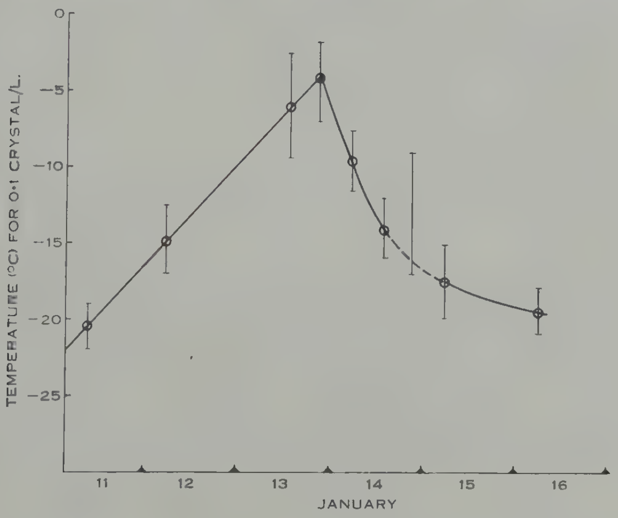


Fig. 3.—The temperatures for 0·1 crystal/l, January 11–16.

The fluctuations in freezing nucleus concentration shown in Figure 2 are quite spectacular. Their relatively regular progress is illustrated in Figure 3 by the example of the period January 11–16, for which some additional experiments were performed. The probable limits of error estimated from the scatter in readings are shown.

IV. Discussion of the Results

(a) *Meteorological Conditions*

The observation, that a concentration of ice crystals as great as 1/l active at −10 °C can occur, is of considerable meteorological significance. It is thought that concentrations of this order of magnitude are sufficient to produce continuous rain from suitable clouds, so that unless such nuclei are purely local their influence on the weather must be great under favourable conditions. It is important therefore to determine whether the nuclei could be of local origin.

On the coastline one might expect the direction of the surface wind to make a marked difference in the numbers of nuclei of terrestrial origin. Days on which the wind was off-shore when the measurements were made were January 11, 18, 19, 28, and 29. Of these days three had below average counts, one average, and one high, showing that there was no consistent effect.

If sea-salt nuclei were also freezing nuclei, a high on-shore wind should lead to high counts. That this was not the primary cause of fluctuations in the freezing nucleus counts is shown by the example of January 13 and 16. On each of these days at the time of the experiment there was a 20–30 knot wind from the sea, yet the temperatures for 0·1/l were −6 and −19 °C respectively.

The properties of the cloud produced in the apparatus could have a possible influence on ice crystal formation. The dew-point, which is a reasonable criterion of the available water vapour, had only a very small range about a mean of 22 °C with the exception of January 11 and 18, when it was low. The count was slightly below average on the 11th and above average on the 18th, which again suggests no significant effect. Upper winds at Carnarvon were variable during January. Frequently a region of off-shore winds occurred between the surface and high altitude on-shore winds. No clear relation to freezing nucleus counts could be found in the characteristics of these winds.

On a wider scale, the pressure distributions did not suggest any cause for the large fluctuations in freezing nucleus counts. Certainly, there was a slightly deeper depression (996 mb) than usual within 500 miles on January 13 when the count was high, but the same was true on January 17 when the count was low. The pressure at Carnarvon itself showed how little the area was influenced by the passage of weather systems to the south, varying only between 1001 and 1010 mb.

The meteorological conditions thus provided no basis for assuming that the nuclei were of local origin.

(b) *Relation to the South African Results*

The average temperature for 1 crystal/l was −22 °C, ignoring peaks, or −19 °C including them. This was considerably warmer than the South African figure of about −33 °C. The different sites, the fact that the measurements were made on the ground rather than in the air, and the slightly different quantities measured could have contributed to this discrepancy. It was possible, however, that much of it was due to inherent differences in the methods of observation.

Of the two significant peaks in the South African results at 20,000 ft, that of January 13 coincided exactly with a West Australian measurement, while that of January 29 was 3 days ahead of the nearest corresponding peak. The degree of coincidence is insufficient to draw any acceptable conclusions.

However, the dates on which the count of airborne particles at 20,000 ft reached a maximum near Pretoria coincided exactly with the dates on which the freezing nucleus concentrations reached a maximum at Carnarvon. If, for some reason, ice formed on these particles but did not produce scintillating crystals, they would be overlooked in a cold box relying on visual detection in a light beam, but not by one using a supercooled sugar solution as a detector. There would then be the very important result that changes in freezing nucleus content in South Africa and Western Australia were simultaneous. There is no evidence for or against the supposition that non-scintillating ice particles may exist, but the possibility is clearly in need of examination.

(*c*) *Relation to Bowen's Meteoritic Dust Theory*

The dates of maximum freezing nucleus counts in January were almost exactly those specified by Bowen's meteoritic dust theory. The probability that this was a coincidence is clearly small in view of the close agreement, the freezing nucleus concentrations having reached maxima close to all the specified dates in January. Carnarvon may have been a more suitable locality for the experiments than others used, or alternatively the equipment used may have been better.

V. Conclusion

The freezing nucleus peaks of January 13 and 21–23 and February 1 appear to have been quite uninfluenced by local weather conditions. Strong support is lent to Bowen's meteoritic dust theory by the appearance of nuclei exactly on or within a day of the dates which it specifies.

The meteorological significance of the abrupt appearance of relatively high concentrations of nuclei active at −10 °C on certain dates, emphasizes the need, pointed out in the preceding paper, for further surveys of this kind.

VI. Acknowledgments

I should like to thank the meteorological staff at Carnarvon for making available local weather data.

COUNTS OF ATMOSPHERIC FREEZING NUCLEI AT PALO ALTO, CALIFORNIA, JANUARY 1956

By R. N. BRACEWELL*

[*Manuscript received June* 29, 1956]

Summary

Measurements of freezing nucleus concentration were made at Palo Alto, California, each day during January 1956. They show maxima on two out of the three dates anticipated from the meteor theory.

I. INTRODUCTION

A cold chamber and cooling tank similar to that shown in Figure 1 of the preceding paper were operated at Palo Alto, California, during January 1956.

Instead of using dry ice as a coolant, the equipment was stored in a deep-freeze unit until required. Measurements were made in precisely the way described in the preceding paper, but were limited in number by the minimum temperature attained and the rate of warming of the cold chamber. Only three measurements were made daily (except on January 20 and 21 when four and five were made, respectively), at temperatures ranging between −20 and −16 °C.

II. RESULTS

The average number of crystals obtained is shown in Figure 1, together with the range of variation. The change in temperature between readings probably contributed to their scatter, but there were insufficient observations to make the necessary correction, except by forming an average over the whole period. Application of this correction made little change to the form of Figure 1. As the South African and Western Australian experience suggests that this procedure may be invalid, the rate of change of number of ice crystals with temperature varying from day to day, the results treated in this way have not been shown.

A peak on January 13 is quite clear in Figure 1, and if the curve is divided into three, representing first, second, and third (or warmest) trials, it shows on each. The peak of January 21 is clearly less significant because of greater scatter and appears on only the first and second experiments.

---

* Stanford University, California, U.S.A.

III. Weather Conditions

Detailed meteorological records for the period were not obtained, but the weather was foggy, overcast, or raining except on January 25 and January 28 to February 5 when it was fine. On the days of particular interest the conditions were :

13 : Overcast all day, no rain or fog.
21 : Raining lightly.
27 : Raining all day.
30 : Clear all day.

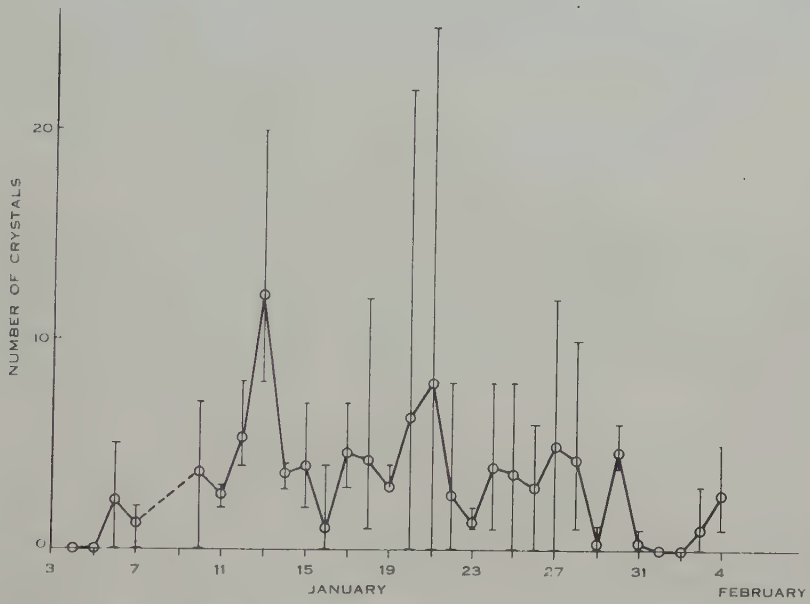


Fig. 1.—Total number of crystals observed in 3 l. at a mean temperature of —18·4 °C at Palo Alto, California.

IV. Discussion

The mean nucleus concentration for the whole sequence of observations was 3·3 per litre at a mean temperature of —18·4 °C, whereas the mean South African temperature for a count of 3 crystals per litre was —34 °C and the mean West Australian temperature —22 °C. Thus the nucleus concentration was apparently very much greater than in South Africa and significantly higher than in Western Australia. Since the South African measurements were made at heights of several thousand feet and the West Australian measurements were made in what was often pure sea air, there is a suggestion that the nuclei counted at Palo Alto included some of terrestrial origin. Of course, the reason for making measurements over Pretoria and at Carnarvon was to avoid possible contamination of the air by nuclei which could mask the meteoritic contribution being sought.

On any one day the consecutive readings showed a scatter consistent with the expected spread due to " patchiness " in the distribution of nuclei plus an erratic component associated with the technique. This scatter, together with the non-meteoritic contribution, considerably reduced the sensitivity of the sequence of observations to any increases in nuclei which might be due to meteoritic dust.

In spite of this, the two principal peaks of Figure 1 fell on January 13 and 21, two of the critical dates in January already established by Bowen.

The author concludes (i) that these measurements support the earlier evidence from rainfall records that certain dates in January have worldwide significance and (ii) that ground level observations with simple equipment in unselected localities appear promising for studying this phenomenon.

SHORT COMMUNICATIONS

THE ELECTRONIC HEAT CAPACITY OF A COPPER–3 PER CENT. ZINC ALLOY*

By J. A. Rayne†‡ and W. R. G. Kemp†

Recent measurements on a copper–1 per cent. cadmium alloy (Rayne 1956) have shown that its electronic heat capacity differs considerably from that of pure copper. For the alloy

$$\gamma=0\cdot799\pm0\cdot005\times10^{-3}\ \mathrm{J\ mole^{-1}\ deg^{-2}},$$

while for the pure metal

$$\gamma=0\cdot686\pm0\cdot005\times10^{-3}\ \mathrm{J\ mole^{-1}\ deg^{-2}}.$$

This difference could be due either to a valence electron effect arising from a steep slope in the density-of-states curve for copper or to the severe lattice distortion produced by the cadmium atoms. To distinguish between these possibilities, the low temperature heat capacity of a copper–3 per cent. zinc alloy was measured. If valence electron effects are predominant, the addition of further divalent atoms would be expected to increase the electronic heat capacity even more.

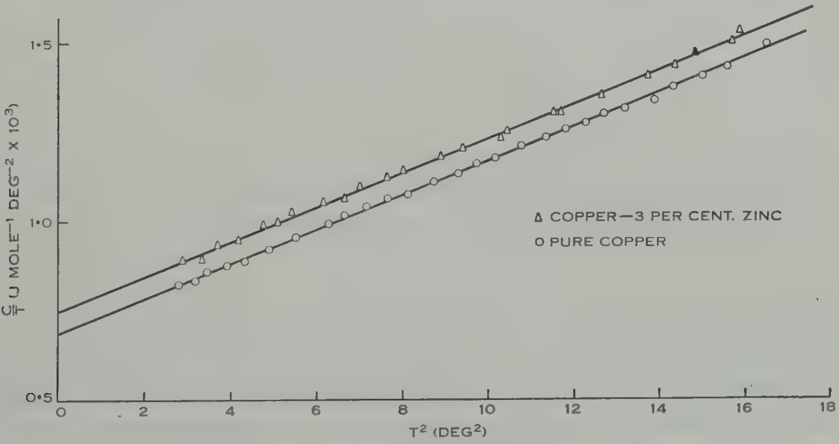


Fig. 1.—Atomic heat of copper-3 per cent. zinc alloy.

The results of the measurements, together with those for pure copper, are given in Figure 1 as a plot of C/T versus T^2.

---

* Manuscript received July 17, 1956.

† Division of Physics, C.S.I.R.O., University Grounds, Chippendale, N.S.W.

‡ Present address: Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania, U.S.A.

A least squares fit of a straight line to the experimental points for the alloy gives

$$\gamma=0\cdot750\pm0\cdot005\times10^{-3}\ \mathrm{J\ mole^{-1}\ deg^{-2}},$$
$$\theta=345\cdot1\pm1\ ^{\circ}\mathrm{K},$$

the quoted uncertainties including both systematic and random errors.

Clearly the change in γ is considerably smaller than that for the cadmium alloy and hence the large electronic heat capacity of the latter must be connected in some way with the large lattice distortion produced by the cadmium atoms. It is of interest to note, however, that the change in γ for the zinc alloy is considerably larger than that expected from the usual theory of such alloys (Mott and Jones 1936) assuming a valence band of standard parabolic form. Although this discrepancy could be due to a non-spherical Fermi surface in copper, such as postulated by Klemens (1954), no definite conclusion can be reached until the magnitude of the effects arising from lattice distortion in the copper-zinc alloys has been established. Further experiments are being carried out to investigate this point.

References

Klemens, P. G. (1954).—*Aust. J. Phys.* 7 : 70.

Mott, N. F., and Jones, H. (1936).—" Properties of Metals and Alloys." (Oxford Univ. Press.)

Rayne, J. A. (1956).—*Aust. J. Phys.* 9 : 189.

AN APPLICATION OF THE DIFFUSION EQUATION TO VISCOUS MOTION WITH A FREE SURFACE*

By J. R. Philip†

The recent development of methods for the numerical solution of the diffusion equation when the diffusivity is concentration-dependent enables certain phenomena, previously not amenable to quantitative analysis or ineffectively described by empirical functions, to be studied more fully. This communication points out an example of concentration-dependent diffusion which does not appear to have previously received attention.

The steady one-dimensional viscous flow of a liquid over any plane surface in a gravitational field may be expressed as

$$q=-D(\mathrm{d}y/\mathrm{d}x-\tan\alpha), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where q is the discharge per unit width normal to the direction of the motion, y is the depth of the fluid, x is the horizontal ordinate, α is the angle the plane

---

* Manuscript received February 6, 1956.

† Division of Plant Industry, C.S.I.R.O., Deniliquin, N.S.W.

makes with the horizontal, and D is a function of y depending on the properties of the surface. Thus for a smooth surface $D=(gy^3 \cos \alpha)/3\nu$, where g is the acceleration due to gravity and ν is the kinematic viscosity of the liquid; for a densely vegetated surface (and depth of flow rather less than the height of the vegetation) with the permeability at height y' equal to K,

$$D=\frac{g \cos \alpha}{\nu}\int_0^y K dy'.$$

For the motions we study here the flow velocity is so small that we may neglect velocity head. If we also make the reasonable, but not exact, assumption that all motion is parallel to the plane surface, (**1**) is valid for unsteady motion. Then, applying the requirement of continuity to (**1**), we obtain

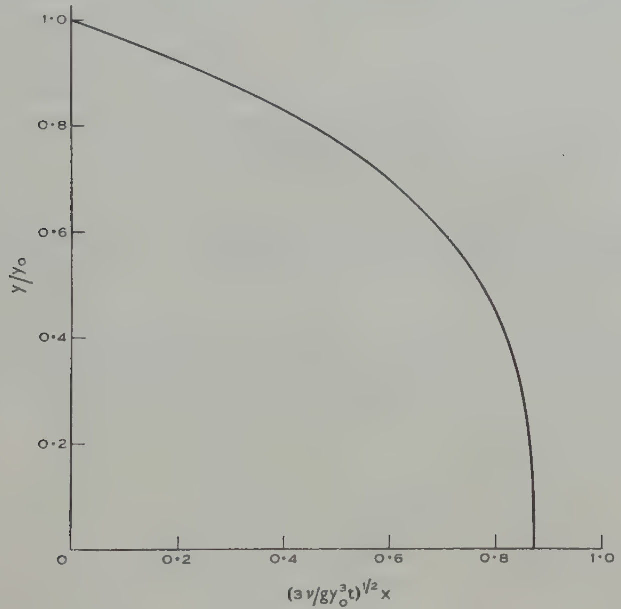


Fig. 1.—The advance of a viscous translatory wave over a smooth dry horizontal surface.

$$\frac{\partial y}{\partial t}=\frac{\partial}{\partial x}\left(D\frac{\partial y}{\partial x}\right)-\tan\ \alpha\frac{\partial D}{\partial x}. \qquad \text{(2)}$$

For the case of a horizontal surface, $\tan \alpha=0$, and (**2**) becomes the well-known diffusion equation with diffusivity concentration-dependent. A rapid method of solving this equation for certain simple but important conditions has been given by Philip (1955). For the case of a sloping plane (**2**) is a generalization of the Fokker-Planck equation. Philip (1957) gives a quick method of solving this equation—for the same simple conditions. These are

$$t=0,\ x\geqslant 0,\ y=y_n,\quad t\geqslant 0,\ x=0,\ y=y_0. \qquad \text{(3)}$$

Conditions (3) correspond to the sudden imposition of a liquid depth y_0 at $x=0$ on a surface originally covered by a depth of liquid y_n. We confine the remainder of this communication to flow over the horizontal surface.

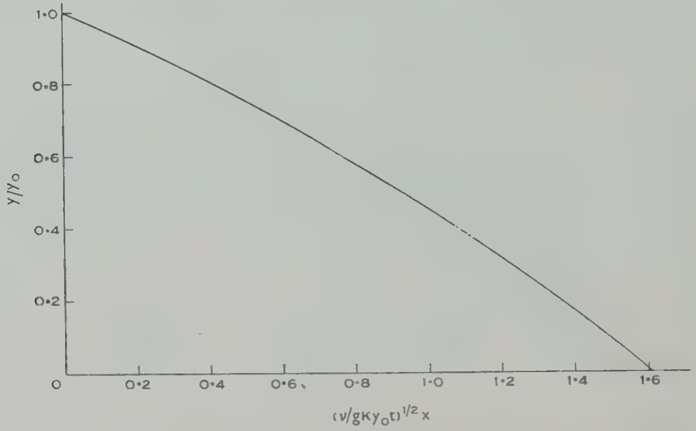


Fig. 2.—The advance of a viscous translatory wave over a vegetated dry horizontal surface with K, the permeability of the vegetation, constant.

An interesting feature of the solution of (2) subject to (3) is that, provided $D(y_n)=0$, the infinite "tail" common in the mathematics of diffusion does not occur, and $y=y_n$ for some finite value of $xt^{-\frac{1}{2}}$. For the important case $y_n=0$, this condition is always met, so that the analysis predicts the advance of a wave front over an initially dry surface. However, for $y_n>0$, $D(y_n)\neq 0$, so that $y\to y_n$ as $x\to\infty$, i.e. the usual infinite tail occurs. Thus a blind observer

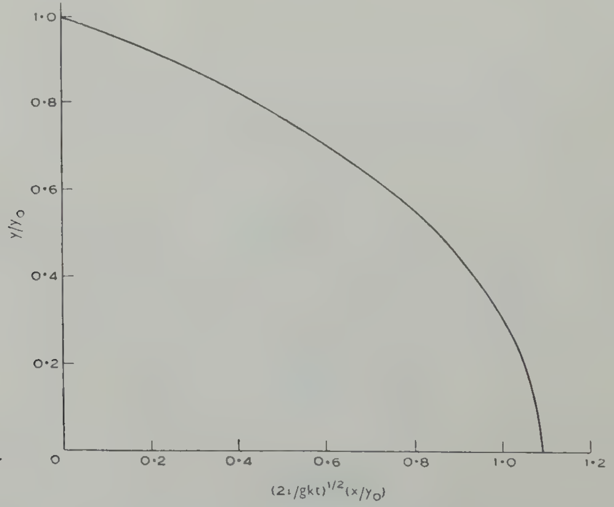


Fig. 3.—The advance of a viscous translatory wave over a vegetated dry horizontal surface with $K=ky$.

at x equipped with very sensitive depth-measuring apparatus would detect the change in boundary conditions *instantaneously* for the case of $y_n>0$, but for $y_n=0$ the disturbance would not be detected until a finite time (proportional to x^2 when $\alpha=0$) after $t=0$.

Solution of three cases of (2) subject to (3) (for $\alpha=0$, $y_n=0$) are given in Figures 1–3. Figure 1 is for flow over a smooth surface. Figure 2 depicts flow over a densely vegetated surface with the permeability of the vegetation, K, constant between depths $y=0$ and $y=y_0$. Figure 3 gives the result of an intermediate case in which the permeability increases linearly with y from zero at the surface to a value ky_0 at $y=y_0$. This case is relevant where the density of vegetation decreases with height. It should be pointed out that the numerical method employed to obtain these solutions is equally applicable to any form of variation of permeability with y.

This communication gives the results of a preliminary study of the problem. It is hoped that it will be possible later to treat these matters in more detail. However, certain conclusions, important in fluid mechanics and hydrology, may be presented at this stage:

(i) The advance of a viscous translatory wave over a dry or initially submerged surface may be analysed by treating the problem as one in diffusion. Since this is a class of motion which has hitherto eluded analysis, this appears to be a significant advance in this sector of fluid mechanics.

(ii) Study of the hydraulics of overland flow, important in storm hydrology and the hydraulics of surface methods of irrigation, has long been retarded by the need for a suitable measure of the hydraulic properties of vegetative cover. The concept of a height-variable permeability seems an acceptable means of representing these properties. Then, if conditions (3) are realized on the surface to be measured and the liquid profile observed at some time t, it is a simple matter to deduce D as a function of y. Then differentiating D with respect to y will give the required $K=K(y)$.

References

PHILIP, J. R. (1955).—*Trans. Faraday Soc.* **51**: 885.

PHILIP, J. R. (1957).—Numerical solution of equations of the diffusion type with diffusivity concentration-dependent. II. *Aust. J. Phys.* **10** (in press).

DOUBLE-LAYER PHENOMENA IN THE E_s REGION*

By J. A. THOMAS†

At Brisbane, ionospheric records are frequently made which show the simultaneous presence of two types of E_s ionization, one at a greater virtual range than the other; these have been classified by McNicol and Gipps (1951) as "sequential" type (E_{ss}), and "constant-height" type (E_{sc}), respectively. Such records occur most frequently in the late afternoon and early evening, particularly in summer, when E_{ss} is in its last stages and E_{sc} is building up to its maximum intensity (Thomas 1956). Very often such "double-layer" records will continue for an hour or longer, until one of the traces (usually the upper one) either simply fades out with the general night-time decrease in ionization, or increases in range as the region moves to one side. Examples of this phenomenon as it occurs on $P'f$ and $P't$ (2·28 Mc/s) records, are shown in Figures 1 and 2.

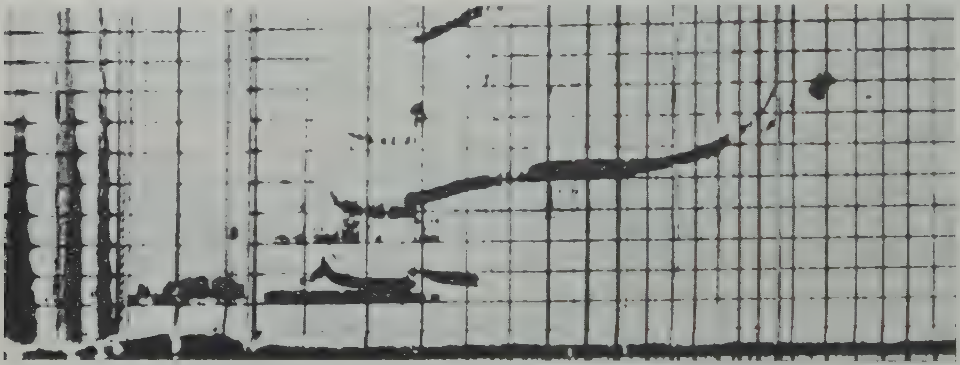


Fig. 1.—$P'f$ record showing echoes from two different E_s regions at 100 and 120 km.

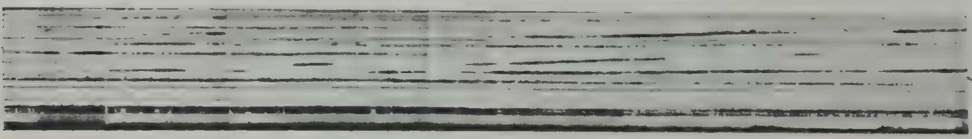


Fig. 2.—$P't$ record (2·28 Mc/s) showing a double E_s layer reflection.

There are a few occasions when, in addition to the two traces due to the double layer, there appears a "stack" of echo traces, each separated by the

---

* Manuscript received August 6, 1956.

† Physics Department, University of Queensland, Brisbane.

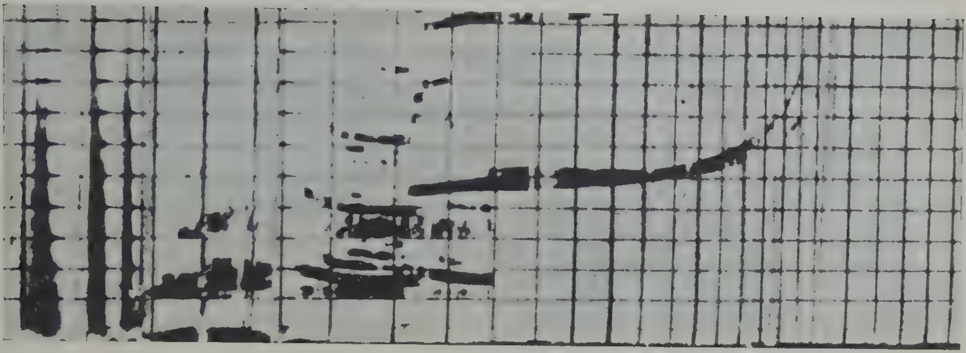


Fig. 3.—$P'f$ record showing multiple-hop echoes arising from a double E_s layer.

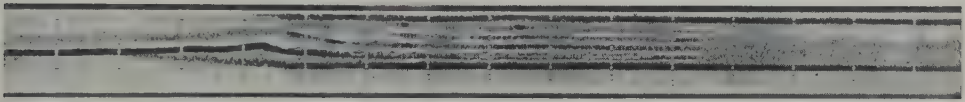


Fig. 4.—$P't$ record showing a "stack" of multiple-hop echoes arising from a double E_s layer.

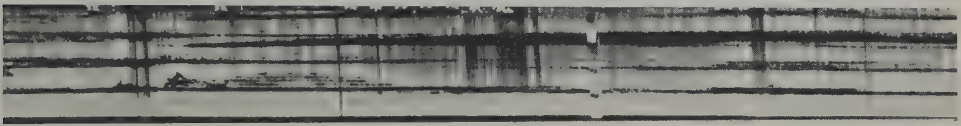


Fig. 5.—$P't$ record showing a "stack" with a curious structure brought about by the movement of slightly tilted regions of a double layer.

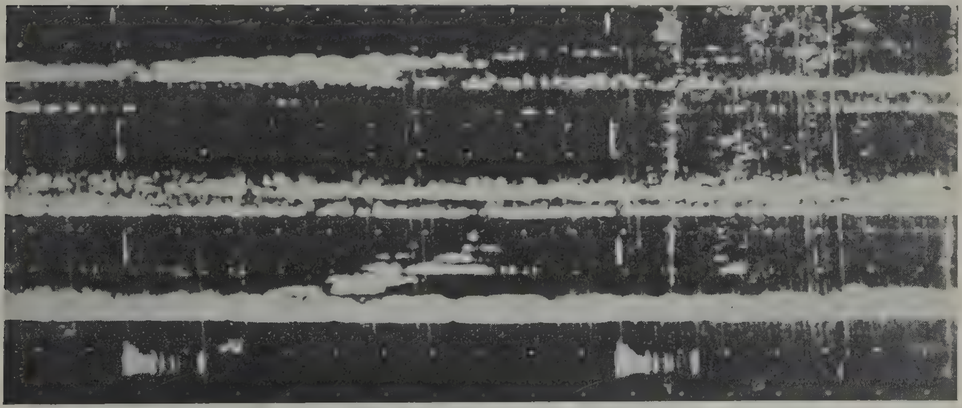
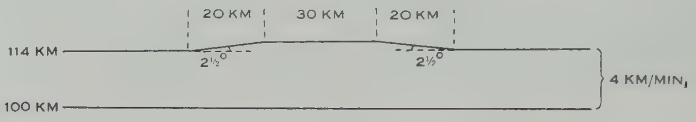


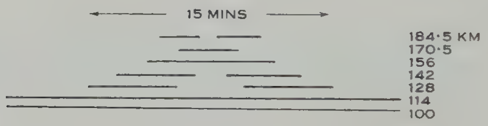
Fig. 6.—$P't$ record showing a "stack" structure due to movement of a tilted region of a double layer.

spacing of the base layers. Figures 3–6 show some examples of such double-layer stacks.

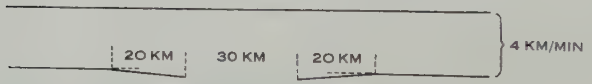
Such echoes can be explained on the basis of multiple hops between the two layers. The structure shown in Figure 4 needs no further explanation; the curious patterns of Figures 5 and 6, however, need further investigation, and an explanation has been devised along the lines given by Baird (1954) in explaining the production of abnormal high multiple F echoes.



(a) LAYER PAIR



(b) ECHO STRUCTURE



(c) ALTERNATIVE LAYER PAIR

Fig. 7.—Structure of hypothetical layer pair, and the echoes to which such a layer pair would give rise.

If we suppose one of the layers to be substantially flat and the other to have regions of small slope (Fig. 7 (*a*)), then horizontal movement of the sloping layer (or both) will give rise to multi-hop echoes of ranges similar to those shown in Figure 5. With a horizontal speed of 4 km/min such a layer system will give rise to the traces sketched in Figure 7 (*b*). This echo structure differs from that in Figure 5, in that in the actual record the lowermost trace is absent for part of the time; this can well be explained if the upper side of the layer is a better reflector than the lower side, which is not uncommon (Thomas 1956). An alternative and more probable explanation is shown in Figure 7 (*c*); such a layer pair gives the correct echo structure including the gap in the lowermost trace.

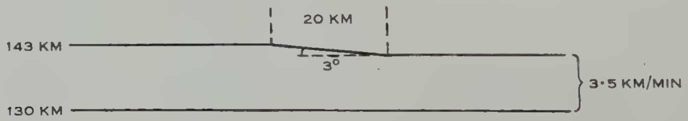


Fig. 8.—Hypothetical layer pair which gives an echo structure very similar to that of Figure 6.

The structure of Figure 6 may be explained by the hypothesis of the movement of a single sloping portion of one of the layers as sketched in Figure 8.

It appears, then, that sloping regions can exist in moving E_s layers. If a layer is regarded as made up of an aggregate of disk-like clouds (Thomas and Burke 1956), then it is logical to expect small sloping regions of the order of those envisaged in the above discussion.

On the other hand, the larger sloping regions necessary to produce high multiple E_s patterns similar to those found for F echoes, are not very likely to occur because of the limited size of E_s clouds. An examination of the records taken over two years revealed only one such pattern and even that was a doubtful example.

This work forms part of the research programme of the Radio Research Board of C.S.I.R.O. and is published by permission of the Board.

References

BAIRD, K. (1954).—*Aust. J. Phys.* **7**: 165.
McNICOL, R. W. E., and GIPPS, G. DE V. (1951).—*J. Geophys. Res.* **56**: 17.
THOMAS, J. A. (1956).—*Aust. J. Phys.* **9**: 228.
THOMAS, J. A., and BURKE, M. J. (1956).—*Aust. J. Phys.* **9**: 440.

INDEX

PAGE

Aerial Smoothing, Two-dimensional, in Radio Astronomy .. 297

Backscatter, High-frequency, Observations at Salisbury, South Australia 454

Ball, F. K.—
  The Theory of Strong Katabatic Winds 373

Beam, Compressed, Bounds for the Collapse Load of a .. 419

Beck, A., Jaeger, J. C., and Newstead, G.—
  The Measurement of the Thermal Conductivities of Rocks by Observations in Boreholes 286

Bigg, E. K.—
  Counts of Atmospheric Freezing Nuclei at Carnarvon, Western Australia, January 1956 561

Bowen, E. G.—
  A Relation between Snow-Cover, Cirrus Cloud, and Freezing Nuclei in the Atmosphere 545
  January Freezing Nucleus Measurements 552

Bowman, G. G.—*See* McNicol, R. W. E. 247

Bracewell, R. N.—
  Counts of Atmospheric Freezing Nuclei at Palo Alto, California, January 1956 566
  Strip Integration in Radio Astronomy 198
  Two-dimensional Aerial Smoothing in Radio Astronomy 297

Brown, H. E.—*See* Thomas, J. A. 159

Bubbles, Accelerated Motion of 19

PAGE

Buchdahl, H. A.—
  Reciprocal Static Solutions of the Equations of the Gravitational Field 13

Bunker, A. F.—
  Measurement of Counter-gradient Heat Flows in the Atmosphere 133

Burke, M. J.—*See* Thomas, J. A. 440

Campbell, J. G.—
  Composition of the ^{7}Li(p,γ) Radiation below the 441 keV Resonance .. 156

Carbon, Scattering of High Energy Electrons by, and the Nuclear Density Distribution 400

Carte, A. E.—*See* Mossop, S. C. 556

Chapman Layer, Behaviour in the Night F_2 Region of the Ionosphere 436

Cloud, Cirrus, A Relation between, and Freezing Nuclei in the Atmosphere 545

Conduction, Heat, in an Infinite Region bounded internally by a Circular Cylinder of a Perfect Conductor 167

Conductivity, Electrical, of Iron, Nickel, Titanium, and Zirconium at Low Temperatures 180

Conductivity, Thermal, of Iron, Nickel, Titanium, and Zirconium at Low Temperatures 180

Conductivity, Thermal, of Rocks by Observations in Boreholes 286

Copper, Heat Capacity of, below 4·2 °K 189

Copper-3 per cent. Zinc Alloy, Electronic Heat Capacity of 569

Corrigendum 296

Cosmic Rays, Diurnal Variation of the Barometer Coefficient at Hobart 505

PAGE

Das, S. C.—
 Statistical Analysis of Australian Pressure Data .. 394
 Statistical Analysis of Australian Temperature Data 542
 The Fitting of a Truncated Log-normal Curve to Daily Rainfall Data 151

Deacon, E. L., Sheppard, P. A., and Webb, E. K.—
 Wind Profiles over the Sea and the Drag at the Sea Surface 511

Diffusion Equation, Application to Viscous Motion with a Free Surface 570

Droplets, Accelerated Motion of 19

Duncan, R. A.—
 Lunar Variations in the Ionosphere 112
 The Behaviour of a Chapman Layer in the Night F_2 Region of the Ionosphere, under the Influence of Gravity, Diffusion, and Attachment 436

Edge, R. D.—
 Neutron Experiments with a Sensitive Szilard-Chalmers Detector 429

Electrons, High Energy, Scattering by Carbon and the Nuclear Density Distribution 400

Electrons, Scattering of 121 keV 54

Electrons, Slow, Formulae for the Mean Losses of Energy in Collisions of, moving in Diatomic Gases 44

Ellyett, C. D., and Keay, C. S. L.—
 Radio Echo Observations of Meteor Activity in the Southern Hemisphere .. 471

Energy, Activation, Calculation of, from Calorimetric Work carried out at a Constant Rate of Heating 425

Expansion, Thermal, of a Grossularite Garnet 403

PAGE

Field, Gravitational, Reciprocal Static Solutions of the Equations of the 13

Freiberger, W.—
 Bounds for the Collapse Load of a Beam Compressed by Three Dies 419

Gani, J.—
 The Condition of Regularity in Simple Markov Chains .. 387

Garnet, Grossularite, Thermal Expansion of a 403

Gas Column, Ionized, On the Dipole Resonant Mode of an 144

Gregory, J. B.—
 Ionospheric Reflections from Heights below the *E* Region 324

Heat Capacity, Electronic, of a Copper-3 per cent. Zinc Alloy 569

Heat Capacity of Copper below 4·2 °K 189

Heat Flows, Counter-gradient, in the Atmosphere 133

Heffernan, K. J.—*See* Mossop, S. C. 556

Heisler, L. H.—*See* Munro, G. H. 343, 359

Hill, G. W.—*See* Pearcy, T. .. 19

Huxley, L. G. H.—
 Formulae for the Mean Losses of Energy in Collisions of Slow Electrons moving in Diatomic Gases 44

Integration, Strip, in Radio Astronomy 198

Ionosphere, Behaviour of a Chapman Layer in the Night F_2 Region of the 436

Ionosphere, Cusp Type Anomalies in Variable Frequency Ionospheric Records 343

Ionosphere, Divergence of Radio Rays in the 359

Ionosphere, Double-layer Phenomena in the E_s Region 574

Ionosphere, High-frequency Backscatter Observations at Salisbury, South Australia .. 454

Ionosphere, Ionospheric Recorders and Sporadic *E* .. 159

PAGE

Ionosphere, Lunar Variations in the 112
Ionosphere, Motion in the Night-time E_s Region at Brisbane . . 440
Ionosphere, Processes controlling Ionization Distribution in the F_2 Region of the 161
Ionosphere, Reflections from Heights below the E Region . . 324
Ionosphere, Sporadic E at Brisbane 228
Ionosphere, " Spread-F " Ionospheric Echoes at Night at Brisbane 247, 272
Iron, Thermal and Electrical Conductivity of, at Low Temperatures 180

Jacklyn, R. M.—
  The Diurnal Variation of the Barometer Coefficient for Cosmic Rays at Hobart . . 505
Jaeger, J. C.—
  Conduction of Heat in an Infinite Region Bounded internally by a Circular Cylinder of a Perfect Conductor 167
  See also Beck, A. 286
Jupiter, 18·3 Mc/s Radiation from 61

Keay, C. S. L.—*See* Ellyett, C. D. 471
Keitel, G. H.—
  On the Dipole Resonant Mode of an Ionized Gas Column 144
Kemp, W. R. G.—*See* Rayne, J. A. 569
Kemp, W. R. G., Klemens, P. G., and White, G. K.—
  Thermal and Electrical Conductivities of Iron, Nickel, Titanium, and Zirconium at Low Temperatures . . 180
Kerr, F. J., and de Vaucouleurs, G.—
  The Masses of the Magellanic Clouds from Radio Observations 90
Klemens, P. G.—*See* Kemp, W. R. G. 180

PAGE

Lithium, Composition of the ^{7}Li(p,γ) Radiation below the 441 keV Resonance 156
Little, A. G.—*See* Mills, B.Y. 84, 218
Love, E. R.—
  Linear Superposition in Viscoelasticity and Theories of Delayed Effects . . 1

McCue, C. G.—
  High-frequency Backscatter Observations at Salisbury, South Australia 454
McNicol, R. W. E., and Webster, H. C.—
  A Study of " Spread-F " Ionospheric Echoes at Night at Brisbane. II. Interpretation of Range Spreading 272
McNicol, R. W. E., Webster, H. C., and Bowman, G. G.—
  A Study of " Spread-F " Ionospheric Echoes at Night at Brisbane. I. Range Spreading (Experimental) . . 247
Magellanic Clouds, Masses of, from Radio Observations . . 90
Markov Chains, the Condition of Regularity in Simple . . 387
Martyn, D. F.—
  Processes controlling Ionization Distribution in the F_2 Region of the Ionosphere 161
Mather, K. B.—
  Neutron Fluxes Underground 147
Meteor Activity in the Southern Hemisphere, Radio Echo Observations of 471
Mills, B. Y., Little, A. G., and Sheridan, K. V.—
  Emission Nebulae as Radio Sources 218
  Radio Emission from Novae and Supernovae 84
Mossop, S. C., Carte, A. E., and Heffernan, K. J.—
  Counts of Atmospheric Freezing Nuclei at Pretoria, January 1956 556

PAGE

Munro, G. H., and Heisler, L. H.—
  Cusp Type Anomalies in Variable Frequency Ionospheric Records 343
  Divergence of Radio Rays in the Ionosphere 359

Nebulae, Emission, as Radio Sources 218
Neutron Experiments with a Sensitive Szilard - Chalmers Detector 429
Neutron Fluxes Underground .. 147
Newstead, G.—*See* Beck, A. .. 286
Nicholas, J. F.—
  Calculation of Activation Energies from Calorimetric Work carried out at a Constant Rate of Heating 425
Nickel, Thermal and Electrical Conductivity of, at Low Temperatures 180
Novae, Radio Emission from .. 84
Nuclear Density Distribution, Scattering of High Energy Electrons by Carbon and the 400
Nuclear Shape Oscillations, A Model of, for γ-Transitions and Electron Excitation .. 407
Nuclei, Atmospheric Freezing, A Relation between Snow Cover, Cirrus Cloud, and .. 545
Nuclei, Atmospheric Freezing, Counts at Carnarvon, Western Australia, January 1956 .. 561
Nuclei, Atmospheric Freezing, Counts at Palo Alto, California, January 1956 .. 566
Nuclei, Atmospheric Freezing, Counts at Pretoria, January 1956 556
Nuclei, Atmospheric Freezing, January Measurements .. 552

Pearcy, T., and Hill, G. W.—
  The Accelerated Motion of Droplets and Bubbles .. 19
Philip, J. R.—
  An Application of the Diffusion Equation to Viscous Motion with a Free Surface 570
Piddington, J. H.—
  Growing Electric Space-charge Waves 31

PAGE

Piddington, J. H., and Trent, G. H.—
  A Survey of Cosmic Radio Emission at 600 Mc/s .. 481
  Cosmic Radio Sources Observed at 600 Mc/s .. 74
Positrons, Scattering of 121 keV 54
Pressure Data, Statistical Analysis of Australian .. 394
Prominences, Solar, Association with Solar Radio Emission .. 315

Radiation, Cosmic Radio, Emission from Novae and Supernovae 84
Radiation, Cosmic Radio, 18·3 Mc/s, from Jupiter .. 61
Radiation, Cosmic Radio, Emission Nebulae as Sources of .. 218
Radiation, Cosmic Radio, Investigation of the Source 06N2A in Gemini 494
Radiation, Cosmic Radio, Sources Observed at 600 Mc/s 74
Radiation, Cosmic Radio, Survey of Emission at 600 Mc/s 481
Radio Astronomy, Strip Integration in 198
Radio Astronomy, Two-dimensional Aerial Smoothing in .. 297
Radio Emission, Solar, Association with Solar Prominences 315
Rainfall Data, the Fitting of a Truncated Log-normal Curve to Daily 151
Rayne, J. A.—
  The Heat Capacity of Copper below 4·2 °K 189
Rayne, J. A., and Kemp, W. R. G.—
  The Electronic Heat Capacity of a Copper-3 per cent. Zinc Alloy 569
Rishbeth, H.—
  An Investigation of the Radio Source 06N2A in Gemini .. 494

Shain, C. A.—
  18·3 Mc/s Radiation from Jupiter 61
Sheppard, P. A.—*See* Deacon, E. L. 511

PAGE

Sheridan, K. V.—*See* Mills, B. Y. 84, 218

Snow Cover, A Relation between, and Freezing Nuclei in the Atmosphere 545

Space-charge Waves, Growing Electric 31

Sun, On the Association of Solar Radio Emission and Solar Prominences 315

Supernovae, Radio Emission from 84

Svenson, A. C.—*See* Thomas, J. A. 159

Szilard - Chalmers Detector, Neutron Experiments with a Sensitive 429

Tassie, L. J.—
- A Model of Nuclear Shape Oscillations for γ-Transitions and Electron Excitation 407
- Scattering of High Energy Electrons by Carbon and the Nuclear Density Distribution 400
- The Scattering of 121 keV Electrons and Positrons .. 54

Temperature Data, Statistical Analysis of Australian .. 542

Thomas, J. A.—
- Double-layer Phenomena in the E_s Region 574
- Sporadic *E* at Brisbane .. 228

Thomas, J. A., and Burke, M. J.—
- Motion in the Night-time E_s Region at Brisbane .. 440

PAGE

Thomas, J. A., Svenson, A. C., and Brown, H. E.—
- Ionospheric Recorders and Sporadic *E* 159

Titanium, Thermal and Electrical Conductivity of, at Low Temperatures 180

Trent, G. H.—*See* Piddington, J. H. 74, 481

de Vaucouleurs, G.—*See* Kerr, F. J. 90

Visco-elasticity, Linear Superposition in, and Theories of Delayed Effects 1

Webb, E. K.—*See* Deacon, E. L. 511

Webster, H. C.—*See* McNicol, R. W. E. 247, 272

White, G. K. — *See* Kemp, W. R. G. 180

Wild, J. P., and Zirin, H.—
- On the Association of Solar Radio Emission and Solar Prominences 315

Wilson, L. G.—
- The Thermal Expansion of a Grossularite Garnet .. 403

Wind Profiles over the Sea and the Drag at the Sea Surface .. 511

Winds, Katabatic, the Theory of Strong 373

Zinc-Copper Alloy, Electronic Heat Capacity of a 569

Zirconium, Thermal and Electrical Conductivity of, at Low Temperatures 180

Zirin, H.—*See* Wild, J. P. .. 315

DATE DUE

JAN 9 '80 PER

FEB 0 3 1984

RETD FEB 0 1 1984

PERIODICALS MUST BE RETURNED
TO PERIODICALS DESK ONLY

DEMCO 38-297

4

3 8198 303 614 505
THE UNIVERSITY OF ILLINOIS AT CHICAGO